Vol. 50, No. 1 January, 1953

Psychological Bulletin

CONTENTS

PUBLISHED BIMONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

The Psychological Bulletin contains evaluative reviews of the literature in various fields of psychology, methodological articles, critical notes, and book reviews. This JOURNAL *does not publish reports of original research or original theoretical articles.*

Editorial communications and manuscripts should be sent to Wayne Dennis, Department of Psychology, Brooklyn College, Brooklyn 10, New York. Books for review should be sent to Edward Girden, at the same address.

Preparation of articles for publication. Authors are strongly advised to follow the general directions given in the "Publication Manual of the American Psychological Association" (*Psychological Bulletin*, 1952, 49 (No. 4, Part 2), 389–446). Special attention should be given to the section on the preparation of the references (pp. 432–440), since this is a particular source of difficulty in long reviews of research literature. *All copy must be double spaced, including the references.*

Reprints. Fifty free offprints are given to contributors of articles, notes, and special reviews. Five copies of the JOURNAL are supplied gratis to the authors of book reviews.

Business communications—including subscriptions, orders of back issues, and changes of address—should be sent to the American Psychological Association, 1333 Sixteenth Street N.W., Washington 6, D. C.

Annual subscription: $8.00 (Foreign $8.50). Single copies, $1.50.

PUBLISHED BIMONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

1333 Sixteenth Street N.W., Washington 6, D.C.

Entered as second class mail matter at the post office at Washington, D.C., under the act of March 3, 1879. Additional entry at the post office at Menasha, Wisconsin. Acceptance for mailing at special rate of postage under the provisions of Sec. 34-40 Par. (D) provided for in Section 538, act of February 28, 1925, authorized August 6, 1947. Printed in U.S.A.

Vol. 50, No. 1 January, 1953

Psychological Bulletin

VARIANCE DESIGNS IN PSYCHOLOGICAL RESEARCH

LEONARD S. KOGAN
Institute of Welfare Research, Community Service Society of New York

About a decade ago Garrett and Zubin (**49**), surveying applications and the potential utility of analysis of variance in psychological research design, pointed out that such techniques had not yet been widely employed. Since that time the number of psychological studies using analysis of variance has become so large that even a listing of titles would be of prohibitive length. Several statistical texts emphasizing variance analysis in psychological research have since appeared (**37, 76, 101**) as well as numerous methodological articles written by psychologists.[1]

The purpose of the present review is to indicate the directions and extent to which analysis of variance designs have been applied in recent psychological research. For the most part references are drawn from papers appearing in the *Journal of Experimental Psychology* and the *Journal of Comparative and Physiological Psychology* during recent years. No attempt has been made, however, to make an exhaustive survey of such applications, special emphasis being placed on papers where problem formulation, design, analysis, and inferences are presented in sufficient detail for the reader to grasp essential methodology and thus implement his understanding of experimental design and analysis over that obtainable from the typical artificialities of statistical texts. It will be assumed that the reader is acquainted with the basic concepts and computational procedures for analysis of variance to the level of Edwards (**37**) and McNemar (**101**). Reference will be made to other readily available sources when necessary.

Problems of terminology present difficulties in any discussion of experimental designs. Psychologists have not been consistent in taking over Fisherian terminology (**44, 45**). While terms such as "factorial design," "latin square," "treatment," "replication," and others have gained widespread usage, terms such as "block," "plot," "varieties," etc., have apparently seemed too agronomic to be commonly used by psychologists. In the material to follow, popular terminology such as that used by Edwards (**37**) will be followed, with some attention being paid to alternative names which have been used.

SINGLE-CLASSIFICATION DESIGNS

Most statistical texts introduce the topic of analysis of variance by describing the partitioning of sums of squares (*SS*) and degrees of freedom (*df*) in the case of the single-classification design. The essence of this design is the presence of a single cri-

[1] Edwards (**37**) presents a bibliography of many of these articles

terion of classification usually represented by several independent groups of *S*s upon whom the same measurements have been taken (**49**). The several groups typically involve the application of different experimental treatments. The usual problem to be answered by the analysis is whether the means of the several groups differ more among themselves than can be attributed to random-sampling variation from a common population. The over-all test of the significance of differences among the means is provided by an F ratio with the numerator derived from the variation of the several means and the denominator based on "pooling" the individual differences within the several groups. Other common names for this design are single-factor design, one-factor design, one-way classification, two-part analysis of variance, single-variable design, simple analysis of variance, between and within analysis, and simple classification of variates.

The most frequently used form of this design is the two-group case where the number of observations in each of the groups may be either equal or unequal. For this case the traditional method of analysis is the t test with $k(=N_1+N_2-2)df$ or the equivalent F ratio with one df in the numerator and k df in the denominator. Because of its widespread familiarity, no illustrations of the two-group case will be presented.

The extension of the single-classification design to more than two groups, despite its simplicity, is not frequently found in the literature. Franklin and Brozek (**47**), investigating the relationship between psychomotor performance and type of practice schedule, made use of a single-classification analysis. Thirty-six *S*s were allocated to six equal groups with comparable means and standard deviations on the basis of performance in "try-out" trials. The groups were then assigned different practice schedules on two psychomotor tests, e.g., three trials a day, three trials a week, etc. The single-classification design was applied in testing the over-all significance of differences among the six group means at specified trials. The analysis of variance, say, at the ninth trial appeared in the form shown in Table 1.

TABLE 1

ANALYSIS OF VARIANCE AT SPECIFIED TRIAL (**47**)

Source of Variation	*df*
Between groups	5
Within groups	30
Total	35

A slight complication appears in the single-classification design when the number of *S*s in each of the several groups is unequal. The computational method of correcting for unequal N's by dividing the total squared for each group by its own N is readily found in all texts. Ammons (**2**) used this so-called unbalanced single-classification design in a study of rotor pursuit performance where eight unequal groups were given different conditions of pre-practice warming-up activity. As in the preceding illustration, Ammons used the design to test the over-all significance of differences among group means at specified trials. A more general example of the single-classification design with unequal N's in the groups is provided by Kelman (**82**) in a study involving the comparison of suggestibility scores for four groups of *S*s classified as Control, Success, Failure, and Ambiguous.

Comment. The single-classification design is the prototype of the classical

experimental dictum of keeping all factors constant but the one being investigated. Reliable inference from this design demands that all conditions other than those which distinguish the several experimental groups be kept comparable from group to group or at least completely randomized among the groups. All variation over and above the differences among means is used to make the estimate of chance fluctuation or experimental error. Whenever possible, the *S*s should be assigned to the several groups in a random manner. Large individual differences or heterogeneity of response among *S*s within the same group enter into the estimate of error and may mask small but real differences among the groups. Failure to reject the null hypothesis, i.e., equality of the several means, is thus often attributable to small size of samples. If, on the other hand, the *S*s of the experiment are kept markedly homogeneous by having them all of the same age, sex, IQ, education, etc., significant differences among groups may be found as a function of experimental variations, but the experimenter will then find it difficult to generalize from his findings to a meaningful population.

It is interesting to note that Franklin and Brozek (**47**) in the study cited above did not actually rely on simple randomization in selecting their six groups of *S*s. Near equality of initial means and standard deviations was "forced" by distributing the *S*s among the groups, not by exact pairing, but by a rough matching of high, moderate, and low scores from group to group. This attempted control of subsequent variation was not, however, taken into account in the analyses of results. It seems probable that the size of the "error variance" might have been reduced (but with the loss of 2 *df*) if the analysis had been carried out for a double classification of data (see below), i.e., by adding another classification on the basis of initial score category. It should be emphasized that the writer is not questioning the conclusions of these investigators but merely using their study to illustrate the point that statistical analysis should in general conform to experimental design if maximum accuracy is to be attained.

The single-classification design is somewhat limited in efficiency because of the characteristic heterogeneity of human and animal material used in psychological research. Although this design furnishes the maximum number of ***error df*** for the given number of observations, the error variance is likely to be relatively large unless the several classes contain a fairly large number of observations. Frequently, a marked reduction in error variance can be gained by a slight modification of design. Perhaps the main usefulness of this design is to serve as an extension of the t test to more than two groups. Not only does the analysis of variance evade the practical problem of carrying out a laborious number of t tests when there are many experimental comparisons to be made, but it can be argued that the over-all F test leads to more dependable inference about possible differences among means. The basis for this argument is the increased reliability or precision of the over-all "error" term as a function of the fact that it is based on more *df* than the error based on any two subgroups. Moreover, such t tests are not independent and "significant" t's tend to be found more frequently than indicated by the chosen level of confidence. Thus even when all samples have actually been chosen at random from the same population, separate t tests often indicate apparent significance of differences. With

six samples, for example, Cochran and Cox (**27**, p. **18**) state that the observed t between the highest and lowest mean will exceed the tabled .05 level about **40** per cent of the time.

Multiple-Classification and Factorial Designs

In the single-classification design it is possible to increase the "sensitivity" of the experiment, i.e., allow the detection of smaller differences among the experimental groups, by using a greater number of cases or by improving the reliability of measuring the dependent variable under consideration. A third technique for increasing the sensitivity of an experiment is by deliberately arranging the design so that known sources of variability can be controlled and separated both from the experimental comparisons and from the estimate of experimental error. One of the major purposes of multiple classification in modern experimental design is to provide methods for minimizing experimental error by the control and isolation of extraneous sources of variation. Perhaps the simplest example of such a controlled arrangement is the method of pairing cases. The reduction of the standard error of difference between the means of paired samples, when the pairing results in significant positive correlation between the samples, illustrates the basic procedure of increasing the sensitivity of an experiment by multiple classification. Here the use of the correlation term in the standard error formula or the equivalent method of analyzing the distribution of differences between paired scores is exactly the same as breaking down the total variation of scores into the three mean squares: between treatments, between pairs, and residual.

The general principle involved in the pairing of cases is to increase the homogeneity of experimental material by employing the arrangement known in experimental agriculture as *randomized blocks* (**45**). This design, as the name implies, consisted originally of the marking out of blocks of land with each experimental treatment then being randomly assigned to plots within each block. Each block is often referred to as a replicate. The resulting yields can then be entered in a two-way table with rows representing the treatments and columns representing the blocks. Analysis of variance separates three sources of variation: treatments, blocks, and error. The psychological analogue to the randomized block is seen to be either the single *S* who receives all experimental treatments in randomized order or a group of comparable *S*s, each of whom is randomly assigned to one of the experimental variations. In animal experiments the block may consist of litter mates, thus allowing the control of variation due to strain, age, weight, etc., while in experiments with humans it is common to form the block on the basis of sex, IQ, socioeconomic level, initial scores on the dependent variable, etc. Many examples of the use of randomized blocks in multiple-classification design will be presented below. In all cases a priori information is used in an attempt to increase the precision of experimental comparisons by removing extraneous sources of variation.

Another basis for multiple classification in experimental design is represented in the so-called factorial design. In this case the investigator is interested in studying the effects of a number of different experimental factors, each of which is varied in two or more ways. The experimental treatments of the factorial design involve all possible combinations of the factors under consideration. In dis-

tinction to the classical rule of holding all but one factor constant, the factorial experiment depends on the simultaneous variation of as many factors or conditions as the experimenter chooses to control. Not only is it usually difficult to keep other relevant conditions constant as demanded by the classical single-factor design, but even if such control were attained the basis of generalization would accordingly be limited to the particular pattern of constancies maintained in the given study. Fisher (**45**) stresses the greater efficiency and comprehensiveness of the factorial study. Efficiency is derived from the fact that several factors may be evaluated with the same precision and by fewer observations than would be the case in carrying out separate studies for each factor. Greater comprehensiveness comes from the possibility of evaluating not only the over-all effects of each of the factors but their interactions as well. A broader basis of inductive generalization is derived from the consideration that each factor is evaluated, not with other factors kept arbitrarily constant, but over the range of variation of the other factors involved in the experiment. Because of these unique properties, psychological experimentation is becoming increasingly characterized by the use of factorial designs, often in combination with the principle of randomized blocks.

The distinction between multiple-classification designs and factorial designs in psychological research (**32**, **34**, **49**) is sometimes difficult to make. Baxter (**6**) has presented a discussion of this distinction. If a given rubric of classification can be taken to represent variation either on a quantitative scale, i.e., different amounts, degrees, or levels of a variable, or on a qualitative continuum, i.e., different categories of a set of experimental conditions or treatments, the particular classification may be called a "factor." On the other hand, if the axis of classification does not represent a quantitative or qualitative variable, e.g., subjects, months, schools, that particular classification would not usually be referred to as a factor in strict parlance. It should be emphasized, however, that this literal conception is not widely adhered to and the term "factorial design" is rather loosely employed, not only by psychologists but also by many statisticians. In any case the analysis of multiple-classification and factorial designs generally involves analogous procedures.

The designs falling in the category of multiple classification are most simply referred to in terms of the number of classifications of the data or in terms of the number of factors involved. Thus one may refer to two-way, three-way, etc. classifications or two-factor, three-factor, etc. designs. Other terms which are sometimes used are complex design, three-part, four-part, etc. analysis of variance or, simply, higher-order classifications.

There are several major subcases in multiple-classification and factorial design. The simplest case is that in which there is but one replication, each subclass containing a single observation. In this case the estimate of experimental error is provided by the highest order interaction term. The second case is the design where the subclasses of the multiple classification or each unique factorial combination contain equal numbers of observations. The third case entails frequencies in the subclasses which are proportionate with the marginal totals. And, finally, there is the complex case where the subclasses contain unequal and disproportionate

numbers of observations. Examples will be provided below for each of these variations in fundamental design.

Double classification with one observation per subclass. Carpenter (**17**) carried out a study of the effect of prolonged visual search, submitting his results as evidence that rate of blinking can be used as a criterion of visual efficiency. Twenty *S*s were engaged in a visual task (Mackworth's Clock Test) where they responded to a specified cue by pressing a key twelve times during each half-hour. The measure analyzed was the number of eyeblinks during a two-hour run. The mean number of blinks per minute in each half-hour was calculated for each *S* and these means were treated as single observations. The analysis appeared as in Table 2.

TABLE 2

MEAN NUMBER OF BLINKS PER MINUTE FOR EACH *S* IN EACH HALF-HOUR (**17**)

Subject	Blink Rate per Minute				
	Half-Hours				Mean
	1	2	3	4	
1	8.4	16.9	16.2	17.2	14.7
2	7.3	14.2	15.4	16.6	13.4
3	10.0	15.6	19.4	19.6	16.2
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
18	12.2	16.9	18.2	19.7	16.8
19	59.1	60.5	42.2	83.7	61.4
20	11.1	21.9	16.9	30.3	20.1
Mean	17.4	21.9	21.1	24.8	21.3

Analysis of Variance

Source of Variation	*df*
Between half-hours	3
Between subjects	19
Residual (error)	57
Total	79

It should be noted in Table 2 that the error estimate is actually based on the interaction between half-hours and *S*s. This example illustrates the general form of the double-classification design where there is one observation in each subclass, but in this case the *S*s cannot be regarded as "randomized blocks" since the columns represent successive periods of time and not a random arrangement of different experimental conditions. Although such a refinement was not apparently necessary in this study to demonstrate the "significant" increase in blinking rate, in some cases it may serve to make the comparison of successive time periods more sensitive if individual variations in time regression are taken out of the "error" term as described in the section on "Repeated Measurements" presented below.

Similar two-way classifications, one axis representing *S*s and the other based on successive periods of time, were used by Siegel and Stuckey (**122**) in a study of the diurnal course of water and food intake in rats. The use of a double-classification design with *S*s operating as randomized blocks is found in a study by Chapanis, Rouse, and Schachter (**18**) of the effects of various kinds of intersensory stimulation on form discrimination at low brightness. Although only three *S*s were used in this latter study, with each *S* receiving a random arrangement of six experimental conditions, the design illustrates how an overwhelming amount of consistent individual differences may be separated from the estimate of experimental error by the use of *S*s as randomized blocks. Another example of a double-classification design with *S*s as one criterion of classification is

provided by Postman (**111**) in an experiment relating the efficiency of recognition of nonsense syllables to number of correct and incorrect items in the recognition tests. Double-classification designs with three *S*s as one axis of classification were also used by Mann and Passey (**100**) in a study of adjustment to the postural vertical as a function of magnitude of tilt and duration of exposure. Although this study was factorial in design (8 durations of exposure time and 6 variations of tilt), the investigators neglected the opportunity of evaluating possible interaction between tilt and exposure time by treating the two factors in separate double-classification analyses.

Double classification with equal numbers of observations per subclass. In this design the double classification is replicated so that there are equal numbers of observations within the subcells. This availability of replication allows the "interaction" term used as "error" in the preceding design to be itself tested against the residual "within cells" mean square. In some studies the experimenter may be particularly interested in possible interaction effects and it is impossible to make a judgment about the possible significance of such effects without some form of replication. Chapanis and Leyzorek (**19**) employed this design in a study on accuracy of visual interpolation. Eleven *S*s were given randomly arranged trials where the task was to estimate the position of stimuli by means of 11 different numerical scales. Standard deviation scores based on 25 estimates with each of two different instruments were computed for each *S* for each scale and the two resulting scores were treated as replications within the subclasses.[2] The form of analysis is indicated in Table 3. Although the investigators chose to consider the two scores within each subclass as simple replications leading to the analysis presented in Table 3, a somewhat more informative analysis might have been made by treating the experiment as a triple-classification design with *S*s as one axis of classification. Since each *S* was tested "randomly" on the same two instruments, it would appear that the instruments could be used as a third axis of classification in the form presented below in Table 4. If this had been done, the total *df* would have been allocated as follows: 10 *df* for *S*s, 10 *df* for scales, 1 *df* for instruments, 100 *df* for interaction between *S*s and scales, 10 *df* for interaction between *S*s and instruments, 10 *df* for interaction between scales and instruments, and 100 *df* for the triple interaction.

Two-factor designs with equal numbers of observations in the subcells. This design involves the investigation of the effects of two factors, each of which is varied over a designated number of levels. Equal numbers of different *S*s are randomly assigned to each of the several factorial combinations. Whereas in the preceding double-classification design with *S*s as one of the axes of classification the differences between the effects of the experimental treatments were associated with intrasubject variation, in this design differences in treatment effects are associated with intersubject variation. The basic estimate of experimental error is derived from differences in response of *S*s subjected to the same experimental conditions. The presence of possible interactions between the two factors

[2] It should be noted that this example consists of an analysis of variance of a set of sample standard deviations. Bartlett (**5**) recommends that the analysis be carried out with a logarithmic transformation of variances in such cases.

TABLE 3

STANDARD DEVIATIONS OF RELATIVE ERRORS OF ESTIMATION FOR EACH SUBJECT AND EACH CONDITION (19)

(The entry in each cell is based on 25 estimates.)

Subject	*Instrument*	*Number Scale*					*Mean*
		1000	2000	...	10000	2.5	
A	1	3.96	2.74	...	3.42	2.74	4.49
	2	3.95	3.36	...	3.19	4.48	
B	1	3.36	2.67	...	3.34	6.62	
	2	4.72	3.86	...	3.66	10.64	4.74
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
K	1	2.73	5.58	...	3.15	2.89	4.28
	2	3.85	3.57	...	2.59	3.10	
Mean*		3.99	4.63	...	3.89	7.32	5.14

Analysis of Variance

Source of Variation	*df*
Between subject means	10
Between scale means	10
Subject-scale interaction	100
Between instruments within cells	121
Total	241

* Mean, Instrument 1 = 5.21; Mean, Instrument 2 = 5.08.

may be evaluated when this design is used. Analytic procedure is the same as for the double-classification design with equal numbers of observations in the subcells. Many examples of this so-called replicated two-factor design were found in the literature.

Kimble and Bilodeau (85) employed a 2×2 factorial design in a motor learning study in which initial and final scores on the Minnesota Rate of Manipulation Test were analyzed as a function of two conditions of work and two conditions of rest, with 24 *S*s in each of the four possible combinations of conditions. Other examples were a 2×3 design with eight *S*s per combination used by Norris and Grant (108) in a study of eyelid conditioning as a function of inhibitory or passive instructions and three conditions of reinforcement; a 2×2 design with six *S*s per cell used by Lawrence and Miller (89) in investigating resistance to extinction as a function of two variations in number of reinforced trials and two amounts of reinforcement; a 4×4 design with five *S*s per cell applied by Grant and Schneider (60) in a study of the magnitude of GSR response during extinction as a function of

four levels of CS intensity during both reinforcement and extinction; a 4×4 design with four *S*s per cell used by Grant and Schneider (**59**) in studying the relation of intensity and frequency of a conditioned eyelid response during extinction to four variations in intensity of CS during reinforcement and extinction; a 3×3 design with ten *S*s per cell by Chernikoff and Brogden (**20**) in a study of the effects upon sensory conditioning of three variations in pretraining treatment and three types of instructions; and a 2×2 design with 20 *S*s per cell used by Grant, Norris, and Boissard (**57**) in studying the change in mean magnitude of eyelid response from pretest to posttest as a function of the presence or absence of dark adaptation and the presence or absence of pseudo-conditioning reinforcement.

Double-classification or two-factor designs with unequal but proportionate numbers of observations in the subclasses. This design differs from the usual two-way classification described above in that the numbers of observations in the subclasses, although not the same, are proportionate for each row and column to the numbers of observations in the marginal totals. For example, a 2×2 table with one row having subclasses containing two and four observations and the second row containing three and six observations would fit this description. The analysis of variance for this design offers no computational difficulties since the *SS* for rows, columns, row by column interaction, and within subclasses are additive to the total *SS*. The only corrections necessary for the unequal entries in the subclasses are the same as those used in the analysis of the single-classification design with unequal numbers of cases in the several groups (**124**). Webb (**130**) used this design in a study of the strength of a food-reinforced response as a function of varying conditions of an irrelevant drive. The irrelevant drive in this case was thirst, and one classification of his data was based on the setting up of four independent groups with different periods of thirst deprivation. Each of these groups contained a total of 18 rats. The second classification was based on differentiating the sex of the *S*. The unequal but proportionate subclass frequencies resulted from the fact that each group consisted of 10 males and 8 females.

In his analysis Webb made the assumption that the within-subclass mean square provided the appropriate estimate of error variance for testing the over-all significance of differences among the four major experimental groups. This was a warranted procedure since in all of the measures analyzed (latency, extinction) the *F* ratios of group-by-sex-interaction mean square to within-subclasses mean square were "nonsignificant." In some studies, however, the experimenter might find that the "interaction" is "significant," and he may then desire to test the intrinsic effect of the main classification under the assumption that the appropriate error term should include a compounding of both interaction variation and within-subclass variation. In the ordinary case where the numbers of entries in each subclass are equal, this is done simply by forming an *F* ratio of main effect mean square to interaction mean square. In the present design, however, where the numbers of observations in the subclasses are unequal but proportionate, Smith (**123**) has recently called attention to a qualification in procedure when the investigator desires to test the significance of a main effect over and above the

variation due to possible interaction. The appropriate method for carrying out this test of significance is somewhat complex and involves setting up an F ratio consisting of multiple terms in both numerator and denominator.

An example of a two-factor design with proportionate subclass N's is also reported by Kelman (**82**). As in Webb's study, the interaction was not found to be significant and no complications developed in testing the main effects.

Double-classification or two-factor designs with disproportionate numbers of observations in the subclasses. The situation sometimes arises in experimentation or investigation where the numbers of observations in each of the subclasses of a multiple-classification design are not only unequal but also disproportionate with the marginal totals. In such cases the simple corrections for unequal subclass frequencies which are applied in single-classification designs or multiple-classification designs with proportionate frequencies are no longer adequate. Such a state of affairs may arise because of various reasons, e.g., failure of Ss to meet appointments, loss of animals, type of investigation, etc. Such designs are referred to as the "nonorthogonal case" because the estimates of variance computed for the several sources of variation are interdependent (**124**). Thus in a 2×2 classification if one were to calculate separately the SS for columns, rows, column-by-row interaction, and residual he would find that these SS would not generally add to the total SS.

In the simplest case only one or two items of data may be missing from some of the cells. The common method for estimating a small number of missing entries and filling out a table was developed by Yates (**132**) and is readily accessible in Snedecor (**124**), Anderson(**3**), and Cochran and Cox (**27**). The general problem of analyzing tables of multiple classification with disproportionate subclass numbers is discussed by Lindquist (**95**) and Johnson (**76**), but in the absence of specific cautions, students referring to McNemar (**101**) and Edwards (**37**) may incorrectly infer that the corrections described for single-classification inequality of frequencies are sufficient. A number of different solutions to the problems of disproportionate frequencies have been proposed, all of which involve approximations based on varying assumptions. Snedecor (**124**) has presented a comprehensive summary of the so-called methods of fitting constants, unweighted means, expected subclass numbers, and weighted squares of means. In these methods it is generally assumed that the usual within-subclasses SS furnish an appropriate estimate of error variance. Tsao (**128**), on the other hand, has derived solutions where this assumption is not made. Two of the basic decisions which the investigator must always make in selecting a solution are whether or not interaction is "significant" and whether or not disproportionality is characteristic of the inferred population.

In employing multiple-classification designs with disproportionate subclass frequencies, some psychologists have taken cognizance of the special methods necessary for this case while other studies have been reported in which no apparent adjustments were made. Bray (**11**) used corrections suggested by Snedecor (**124**) in analyzing conformity scores in an autokinetic situation of 2×2 design, where unequal numbers of Ss were classified according to racial attitude and whether or not the confederate was a member of a specified

race. Porter, Stone, and Eriksen (**110**) also used Snedecor's methods in analyzing 2×3 and 2×9 designs in a study where maze error scores were being compared for rats given electroconvulsive shocks in late infancy and control litter mates. Some studies in which the analysis apparently failed to take adequate account of disproportionate subclass frequencies were a 2×2 design by Jenkins and Postman (**75**), a 3×3 design by Postman and Jenkins (**112**), a 2×10 design by Hunt, Schlosberg, Solomon, and Stellar (**72**), and 2×2 and 3×2×2 designs by Citron, Chein, and Harding (**24**).

In some experiments the investigator has sought to evade the problem of disproportionate subclass frequencies by ignoring the individual observations in the cells and analyzing the data as if there were no replication, i.e., analyzing subclass means as if they were single observations. In general this procedure cannot be rigorously defended, especially when the frequencies are markedly dissimilar, since such means are differentially reliable and nonorthogonality is still inherent in the data.

Triple-classification designs with subjects as a criterion of classification. In many essentially two-factor designs a third axis of classification is provided by the fact that each *S* undergoes all of the experimental variations or conditions, frequently in random order. Because "between subjects" is considered a major source of variation in such designs, there is no "within subclasses" estimate of error and the basic estimate of experimental error is provided by the triple- or second-order interaction mean square. Such a design was used by Solomon (**125**) in a study of the effect of effort upon distance discrimination, where ten rats went through four successive experimental sessions, alternately running a maze with and without a load over a period of eight days. Analysis of the ordinal number of the side alley first entered during each session followed the form of Table 4. In this case the *S*s cannot be regarded as "randomized blocks" since each received the same sequence of experimental variations.

By way of didactic comment about Solomon's analysis, the 9 *df* for rats might have been separated into one *df* for sex and 8 *df* for rats within sex groups. Possible sex difference might then have been evaluated by means of an *F* ratio derived from these two sources of variation. Furthermore, it should be noted that days (not analyzed) are confounded with sessions. Finally, the comparison of effort levels is also confounded with days since performance under the condition of "load" as a whole took place one day later than performance without the load. In this case, however, the general temporal trend was to enter a more remote alley and this was an opposing trend to the tendency exhibited under load. One would thus predict that the apparent difference between effort levels might have been even greater, had the two experimental conditions been randomized for each *S*.

Littman (**97**) used a similar 4×4 ×11 design with two groups of 11 *S*s in a study of the generalization of a conditioned GSR to tones other than the original CS. Other applications of this design were made by Black (**10**) in a 5×2×25 study of intensity of oral responses to two types of messages under five levels of intensity, and by Beebe-Center, Black, Hoffman, and Wade (**7**) in a 3×12×9 investigation of per diem consumption as a measure of preference in the rat.

Trifactorial designs with one observation per subclass. The analysis of the triple-factor design with one ob-

TABLE 4

ORDINAL NUMBER OF SIDE ALLEY FIRST ENTERED BY RATS DURING THE EIGHT TEST SESSIONS (125)

		Session—Pairs									
		1		2		3		4			
		Session (day)									
		70	71	72	73	74	75	76	77		
		Level of Effort								Mean	
	Rats	*Normal*	*Load*	*Normal*	*Load*	*Normal*	*Load*	*Normal*	*Load*	*Normal*	*Load*
♂	1	3	3	3	2	4	4	4	3	3.5	3.0
	2	4	3	3	3	5	5	4	5	4.0	4.0
	3	4	3	4	2	5	4	4	4	4.3	3.3
	4	3	1	3	2	4	2	4	2	3.5	1.8
	5	3	3	4	4	4	4	4	3	3.8	3.5
♀	1	4	1	4	3	4	3	3	3	3.8	2.5
	2	2	1	2	1	1	1	2	1	1.8	1.0
	3	3	3	4	3	4	3	4	4	3.8	3.3
	4	2	2	4	2	4	3	3	1	3.3	2.0
	5	3	2	5	4	3	2	4	3	3.8	2.8
Day Means		3.1	2.2	3.6	2.6	3.8	3.1	3.6	2.9	3.5	2.7
Session-Pairs		2.7		3.1		3.5		3.3			

Analysis of Variance

Source of Variation	*df*
Rats (R)	9
Effort levels (E)	1
Sessions (S)	3
R×E	9
R×S	27
E×S	3
Error (R×E×S)	27
Total	79

servation per subclass is analogous to that shown in Table 4, with the replacement of rats, i.e., subjects, by the third factor. Helson (70) utilized this design in analyzing a 2×4×10 factorial experiment where time errors with handwheels were classified according to wheel diameter, amount of friction, and speed of turning. Actually, in this study subcell values were averages for different groups of six *S*s each.

Trifactorial designs with replications. The presence of equal numbers of observations in the subclasses of a three-factor design affords a within-cells residual which can be used in testing the significance of the triple interaction. Wilson (**131**) used this design in a study of the frequency of remote associations at recall for rote learning. His application, using three *S*s in each combination of a 4×4×3 design, yielded an analysis as in Table 5.

Gebhard (**50**) used a similar design in a 2×2×2 study investigating attractiveness rankings of tasks classified according to experience (success-failure), expectation of task difficulty, and strength of need. Other investigators using this design were Grant (**55**) in a 2×2×2 factorial study of responses to a card sorting task; Grant, Hornseth, and Hake (**65**) in a 2×2×2 study of the influence of intertrial interval on the Humphreys' effect with verbal responses; and Grant and Mote (**63**) in another 2×2×2 study of the effects of brief flashes of light upon dark adaptation. Lawrence (**88**) reported a study involving a 2×2×2 design in which certain comparisons were confounded because of the nature of the experimental design. A study by Conklin (**28**) which apparently involved a three-factor design in an investigation of the effects of temperature, duration of session, and adaptation on skin resistance presents an allocation of *df* which is difficult to reconstruct.

The problem of disproportionate subclass frequencies with tables of multiple classification again arises in this design. In a 2×2×5 design used by Bendig and Braun (**8**) for studying maze behavior, adjustments were made both for missing cell entries and for differing subgroup *N*'s according to suggestions by Snedecor (**124**), Anderson (**3**), and Schoenfeld (**121**). However, in a 2×4×2 study by Newman and Scheffler (**107**) concerned with sex differences in emotional reaction to the news, where sex, educational level, and type of newspaper were treated as major sources of variation, there is no evidence that account was taken of the markedly disproportionate frequencies in the subclasses.

TABLE 5

Frequency of Remote Associations at Recall (**131**)

Interval between Learning & Recall (mins.)	*Spacing between Trials (secs.)*	*Degree of Learning (% of perfect anticipation)*			
		50	75	100	200
	6	—*	—	—	—
0	30	—	—	—	—
	60	—	—	—	—
	6	—	—	—	—
2	30	—	—	—	—
	60	—	—	—	—
	6	—	—	—	—
5	30	—	—	—	—
	60	—	—	—	—
	6	—	—	—	—
20	30	—	—	—	—
	60	—	—	—	—

Analysis of Variance

Source of Variation	*df*
Degree of learning (D)	3
Intervals following learning (I)	3
Conditions of spacing (C)	2
D×I	9
D×C	6
I×C	6
D×I×C	18
Within cells	96
Total	143

* 3 *S*s per cell; data not provided.

Quadruple and higher classification designs. These designs represent

further elaboration of the principles already described. In some studies all of the classifications can be regarded as factors while in other studies one of the classifications of the data depends upon the fact that each S undergoes every variation of experimental combinations. Quadruple-classification $4\times2\times2\times2$ designs were used by Preston, Spiers, and Trasoff (**116**) in a level of aspiration study. Grant, Hornseth, and Hake (**61**) applied a $5\times2\times4\times40$ design, with *S*s as one criterion of classification, in a study of sensitization of the beta-response to visual stimuli. Littman (**98**) used a $3\times2\times2\times6$ design in a latent learning experiment, while Horowitz (**71**) applied several $10\times4\times4\times2$ designs with *S*s as one classification in a study of visual acuity. Child and Grosslight (**23**) made use of a $3\times2\times2\times2$ factorial design in a study of substitute activity with the added complication of breaking down one of the factors into a major and minor subclassification. A five-way $7\times10\times2\times4\times14$ classification was used by Kuntz and Sleight (**87**) in a study of legibility of numerals as a function of height/width ratio, type of numeral, background, and brightness. The highest number of criteria of classification found in the literature surveyed was applied by Licklider, Bindra, and Pollack (**94**) in a study comparing the intelligibility of normal and "square" speech. Two "talkers" and two "listeners" furnished two of the major criteria of classification in a $2\times5\times2\times2\times3\times10$ design. The authors present an interesting argument for the rationale of generalizing from such a small number of *S*s.

Comment. This section has dealt with the possibilities for increasing the precision and scope of experiments by use of randomized blocks and factorial design. In planning an experiment involving the comparison of the effects of several experimental variations, the investigator must always decide whether to use the same, matched, or different *S*s for the various treatments or treatment combinations. If the same or matched *S*s undergo all treatments in randomized order as in the usual factorial design, it is often possible to increase the precision of experimental comparisons by removing variation associated with over-all differences among such "blocks." Assuming that the total number of observations is the same, such an advantage must be weighed against the broader basis for generalization which is derived from the use of a larger number of randomly assigned *S*s. In experiments where naïveté is essential for *S*s undergoing a given treatment it is obvious that the design should contain different *S*s in each of the subclasses. Similarly, wide individual variations in practice or fatigue effects in the design where each S undergoes all experimental combinations would tend to result in marked interactions between *S*s and treatments, thus tending to obscure differences in the main effects of the several factors. If temporal variation is itself a main subject of investigation little would be gained from the conclusion that *S*s show consistent temporal trends when each S has undergone several experimental treatments in randomized order. The following section on "Repeated Measurements" will present some common useful designs when temporal trend is a main topic of study.

Factorial designs, often involving a fairly sizable number of factors, have become very prominent in recent psychological research.[3] In the main,

[3] Edwards and Horst (**40**) have facilitated the computations involved in higher-order multiple-classification designs by furnishing a method for the direct calculation of second-order and higher interaction *SS*.

such designs have been a boon to experimental methods because they allow the systematic, economical exploration of the effects of a number of different factors as well as possible interactions among the factors. Programs of research, sequentially investigating the effects of varying one experimental factor at a time, such as characterized the field of learning in the past, can be immeasurably hastened and increased in generality by the application of factorial designs. On the other hand, there seems to be a tendency on the part of some experimenters to sacrifice considerations of sample size, representativeness of samples, and both reliability and validity of measurement in their enthusiastic endeavor to test large numbers of hypotheses by means of factorially designed experiments. At the extreme, for example, a complex factorial study providing many *df*'s for making many tests of significance might be carried out for a single *S*. Multiple observations could be secured for each subclass by measuring the dependent variable several times for each treatment combination. The precision of such an experiment might be very high and conclusions valid for the unique *S*, but who would attempt to generalize from the results, whether null hypotheses were rejected or not? The writer discovered no instance of the use of a single *S* in factorial design, but many investigators have reported experiments in which broad inferences were drawn from less than a half dozen *S*s.

Although, in principle, there is no limitation on the number of experimental factors which may be involved, difficulties frequently arise in the interpretation of complex factorial designs. The number of treatment combinations increases very rapidly and often limitations in apparatus or other circumstances cause a large-scale experiment to stretch out over a considerable period of time. The classic example of a fairly elaborate factorial experiment in psychological research is that of Crutchfield (**32**, **33**). In this study the topic of investigation was string-pulling in rats as a function of five factors, each varied over three levels. A single animal was assigned to each of the 243 treatment combinations. Complete analysis would yield a list of 31 mean squares: 5 main effects, 10 two-factor interactions, 10 three-factor interactions, 5 four-factor interactions, and 1 five-factor interaction. In such a case it is generally assumed that interactions involving three or more factors can be "pooled" to provide an adequate estimate of experimental error. Fisher (**44**) describes this procedure of dispensing with absolute replication in estimating error as the method of "hidden replication" and points out the possibilities of loss of precision in tests of significance when high-order interactions are not really negligible.

Whether high-order interactions can ordinarily be assumed to be unimportant in psychological research is problematic, but it is certain that they cannot be evaluated when there is no replication. Furthermore, even with replication they frequently present puzzling problems of interpretation to the experimenter and, since large psychological studies are rarely repeated, there is little opportunity to compare their consistency over a series of experiments.

Psychologists in general have not paid much attention to the practical and experimental advantages of "confounding" in the planning of experiments. Confounding in this connection refers to the deliberate arrangement of the experiment so that certain mean squares represent the effects of more than one known source of variation. Experimenters often go to great length to avoid the

possibility of confounding experimental factors, sometimes to the considerable enlargement of their studies, even when previous studies have fairly well demonstrated that the factors concerned do not interact. Of similar character is the practice of running all possible combinations in a factorial experiment and then combining high-order interactions to estimate experimental error. The basic principle of deliberate confounding is to use "incomplete blocks," i.e., blocks within which all treatment combinations do not occur (**45**). In general, the purpose of such confounding is to increase the precision of selected experimental comparisons while sacrificing the possibility of evaluating other comparisons, e.g., high-order interactions.

A simple illustration will clarify the basic idea in deliberate confounding. Let us suppose that we have a three-factor experiment, each factor being varied over two levels. Representing factors by letters and levels by subscripts the eight possible combinations may be separated into two subgroups: (*a*) $A_1B_1C_1$; $A_1B_2C_2$; $A_2B_1C_2$; $A_2B_2C_1$ and (*b*) $A_1B_1C_2$; $A_1B_2C_1$; $A_2B_1C_1$; $A_2B_2C_2$. In the usual complete factorial experiment where each *S* serves as a block, every *S* would undergo all eight experimental treatments. Let us, however, modify the design so that five *Ss* undergo all the combinations listed after (*a*) while five other *Ss* undergo those listed after (*b*). Each *S* would then represent an incomplete block. The resulting analysis of variance would then allot 1 *df* each to *A*, *B*, *C*, *AB*, *AC*, and *BC;* 9 *df* to *Ss*; and 24 *df* to the error estimate.

The single-factor and two-factor effects are not influenced by differences among *Ss* (blocks) while the three-factor interaction *ABC* is completely confounded with these differences. In the usual "complete" experiment this latter interaction would have been estimated from the difference of (*a*) and (*b*) above. In such a case the decision to employ the confounded design might be based upon the fact that each *S* is available for only half of the experimental sessions, a desire to avoid fatigue or boredom on the part of *Ss*, or any other reason which might justify halving the experimental period for each *S*. The aim in this particular design is to reduce the error variance used to test the significance of the main effects and two-factor interactions by sacrificing the second-order interaction.

The reader should not conclude that confounding is possible only when individual *Ss* serve as blocks. Confounded designs may at times be fruitfully employed when each *S* undergoes only one experimental combination. Nor are such designs limited to the confounding of high-order interactions. Baxter (**6**) has discussed various possibilities for increased precision in experimental research through the use of confounding. In the main, however, the most comprehensive presentations of designs involving deliberate confounding are found in sources dealing primarily with experimental agriculture (**27, 45, 83, 124**). The following are some hypothetical examples of situations where the experimenter might consider the possible advantages of confounding by means of "incomplete" blocks:

1. The *Ss* fall into homogeneous groups, e.g., by sex, age, family, IQ, school, visual acuity, etc., but there are insufficient *Ss* in each group to allow carrying out all treatment combinations. A common example is the limited size of groups of litter mates.

2. Several experimenters might simultaneously handle portions of the entire program, thus speeding up completion of

the experiment. General differences in response of *S*s due to the experimenters could be removed by confounding the blocks (experimenters) with unimportant interactions. The same principle might be applied when different machines are used to present experimental stimuli or in experiments which involve the use of confederates.

3. In some experiments time and space considerations might determine the separation of blocks. Several different experimental rooms may be involved or relevant environmental conditions may vary from day to day. Treatment combinations belonging to a block could be compared with greater precision than would be possible if such sources of heterogeneity were ignored.

In all of these cases it is presumed that there is significant block-to-block variation with respect to the dependent variable being studied.

Repeated-Measurements Designs

In many studies the experimenter is especially interested in analyzing successive changes in measures obtained repetitively from one or more groups of *S*s. Some of the types of investigation in which the problem of repeated measurements arises are (*a*) examination of learning and extinction data; (*b*) studies of dark adaptation; (*c*) investigations of performance and fatigue; (*d*) analysis of sequential measures of physiological or sensory-motor functions for varying treatment groups. The repeated-measurement situation is so common in psychological research that Edwards (**37**) devotes an entire chapter to this topic. Several articles have been largely devoted to this type of design (**1**, **86**, **96**).

Single group with repeated measurements. If there is but a single group of *S*s, the investigator may be primarily interested in determining whether the group in general shows a significant trend during the successive trials or periods. The simplest method of analysis is a double-classification design where rows represent different *S*s and columns represent successive trials. If the *F* test for trials mean square over the *S*s × trials interaction mean square is "significant," one concludes that there is nonrandom trial-by-trial variation, i.e., trial means are not the same (see Table 2). Such an analysis, however, does not indicate whether the trial means follow a regular linear or curvilinear trend. In order to "test" for the presence of a consistent trend, one must fit curves to the data, taking into account both individual and

TABLE 6

Mean Magnitude (Mm) of CRs Averaged for Successive Five-Trial Blocks During First Day Reinforcement Trials*
(**108**)

Subject	*Successive Five-Trial Blocks*				
	1–5	6–10	. . .	36–40	41–45
1	—	—	—	—	—
2	—	—	—	—	—
3	—	—	—	—	—
4	—	—	—	—	—
5	—	—	—	—	—
6	—	—	—	—	—
7	—	—	—	—	—
8	—	—	—	—	—
Mean	0.00	0.00	. . .	.67	.37

Analysis of Variance

Source of Variation	*df*
Group slope	1
Between individual means	7
Between individual slopes	7
Individual deviations from linearity	56
Total	71

* Data in body of table not provided.

group regressions upon the time scale. One of the methods suggested by Alexander (**1**) was applied by Norris and Grant (**108**) in a study of eyelid conditioning to test the statistical significance of group slope in a design involving nine successive five-trial blocks for a group of eight *S*s. The measure analyzed was mean magnitude of a CR and the analysis appeared as in Table 6. Alexander (**1**) points out that in some cases apparent "significance" of group slope may be attributable to wide variations in individual slopes.

TABLE 7

LOG LATENCIES FOR ACQUISITION OF THE RUNNING RESPONSE* (**93**)

Rats	Acquisition Trials							
	1	2	3	4	5	6	7	8
Group A: Running First, Bar-Pressing Second								
1	—	—	—	—	—	—	—	—
2	—	—	—	—	—	—	—	—
.								
.								
.								
23	—	—	—	—	—	—	—	—
24	—	—	—	—	—	—	—	—
Group B: Bar-Pressing First, Running Second								
25	—	—	—	—	—	—	—	—
26	—	—	—	—	—	—	—	—
.								
.								
.								
47	—	—	—	—	—	—	—	—
48	—	—	—	—	—	—	—	—

Analysis of Variance

Source of Variation	*df*
Trials (T)	7
Between groups (G)	1
Between *S*s in same group (S)	46
Interaction: T×G	7
Interaction: T×S	322
Total	383

* Data not provided.

Independent groups with repeated measurements. A more complex case with respect to repeated measurements occurs when several independent "treatment" or "methods" groups are involved and the investigator wishes to compare the trends exhibited by the several groups. If the assumption, among others, is made that individual regressions are parallel, analysis is readily made in terms of a double-classification design with the between-*S*s variation being subclassified into a between-treatments source of variation and a between-*S*s within-treatment-groups source of variation (**37, 86, 96**). This procedure of decomposing composite classifications or variables "nestled" within other variables will frequently prove to be valuable in the complete analysis of many experimental designs (cf. the "split plot" design described in detail by Cochran and Cox [**27**]). This form of analysis was used by Liberman (**93**) in analyzing 8-trial acquisition and extinction trends for two groups of 24 rats in a study of transfer effects. The variance breakdown is shown in Table 7.

Similar analyses of repeated measurements for independent groups were carried out by Furchtgott (**48**) in a study of maze swimming for three groups of rats exposed to different levels of X-irradiation, and by Bernberg (**9**) in comparing the effects of shock and narcosis upon maze-learning ability in young rats.

The analysis of repeated measurements for independent groups takes on a more complex form if possible individual and group variations in linear regression are taken into account. Alexander (**1**) provides an

analytic procedure for such a trend analysis. His suggestions were applied by Grant, Riopelle, and Hake (**64**) in comparing extinction trends for three groups of 15 *S*s, each group being given a different reinforcement pattern for an eyelid CR. A trend analysis of CR magnitude scores was carried out for five successive blocks of trials with the analysis taking the form shown in Table 8.

Similar analyses were found in other studies by Grant, Hake, and Schneider (**58**) and Grant and Norris (**56**).

Repeated measurements in multiple-classification designs. Many variations of the repeated-measurements design may occur. The within-subclasses mean square of a factorially designed experiment may be based on successive measurements of the same *S*s, or several *S*s within the same subclass may be repetitively measured. In a sense, all designs where the same *S*s undergo several experimental variations are applications of the "repeated-measurements" principle, although the term has generally been used to refer to the case where the effect of the same treatment is measured successively over a period of time. Whereas in the usual factorial experiment with *S*s as one axis of classification the *S*s undergo the several experimental combinations in randomized order, it may sometimes be necessary to have all *S*s undergo the same sequence of treatments. In other experiments the *S*s may undergo the treatments in differing orders, but the investigator may wish to eliminate a general temporal effect, e.g., transfer, fatigue, practice effects, from his estimate of experimental error. An example of a single group experiment where all of the *S*s underwent the several treatments in the same order is provided by Bruner, Postman, and Mosteller (**13**). Nine-

TABLE 8

MEAN MAGNITUDE OF CONDITIONED EYELID RESPONSES FOR SUCCESSIVE FIVE-TRIAL BLOCKS DURING EXTINCTION (**64**)

Subject	*Successive Five-Trial Blocks after First 5 Trials* 2	3	4	5	6
	Single Alternation Group				
1	—				—
2	—				—
.	.				.
.	.				.
.	.				.
14	—				—
15	—				—
	Double Alternation Group				
16	—				—
17	—				—
.	.				.
.	.				.
.	.				.
29	—				—
30	—				—
	100% Reinforcement Group				
31	—				—
32	—				—
.	.				.
.	.				.
.	.				.
44	—				—
45	—				—

Analysis of Variance

Source of Variation	*df*
Over-all slope	1
Over-all deviations from linearity	3
Between group means	2
Between group slopes	2
Group deviations from estimate	6
Between individual means	42
Between individual slopes	42
Individual deviations from estimate	126
Total	224

TABLE 9

Reversals per Successive One-Minute Interval of the Schroeder Staircase under Three Instructions (13)

Instructions	*Minutes*	*Subjects*						
		1	2	...	...	...	18	19
"Alternate" M=47.4	1	120	22				42	30
	•							
	•							
	•							
	10	129	11				35	32
"Hold" M=11.5	11	19	10				7	5
	•							
	•							
	•							
	20	4	2				1	9
"Natural" M=21.6	21	54	20				18	12
	•							
	•							
	•							
	30	25	10	...	...	...	20	10

Analysis of Variance

Source of Variation	*df*
Subjects (Su)	18
Set (St)	2
Interaction (Su×St)	36
Time-sequence regression	57
Residual sampling variance	456
Total	569

teen *Ss* were given the task of reversing the Schroeder staircase for successive ten-minute periods under three sets of instructions. Reversals per successive one-minute interval were analyzed with the time-sequence regression lines for the individual *Ss* being taken out of the total variation of scores as a systematic source of variation. The analysis of variance appeared as in Table 9.

Another complex example, where repeated measurements were analyzed, is found in a study by Lawrence and Miller (**89**). Their investigation involved a 2×2 factorial design in which individual and group linear regression lines (**44**) were compared for groups of *Ss*, doubly classified according to number of reinforced trials and amount of reward. This study appears to be fairly unique in that curvilinear regression lines were also examined.

Comment. Experiments involving the repetitive measurement of a criterion variable for one or more groups of *Ss* are characteristic of many areas

of psychological research. Similar experiments are not commonly found in experimental agriculture and uses of variance analysis for repeated-measurements designs represent special adaptations made by psychologists. Perhaps the most uniquely psychological of these applications is the situation where the investigator is particularly interested in the comparison of trends, e.g., learning curves, for several independent groups subjected to different experimental conditions. Traditional methods for comparing such trends were largely limited to comparisons of experimental groups either at specified points of the experiment (see Table 1) or with respect to increment or decrement over specified periods (**1**). Frequently, the experimenter simply compared the over-all means of the several groups, thus completely neglecting the configurations of successive trial means. The analytic methods cited in this section have proven of practical utility for the comparison of group trends, but in general they should be applied with caution since successive measures taken on the same *S*s can hardly be regarded as either randomly distributed or independent. There is need for further theoretical work on the topic of repeated measurements, preferably with the aid of mathematical statisticians.

Despite the fact that repeated measurements often appear to exhibit nonlinear trends the writer found that few investigators go to the trouble of fitting curvilinear functions in carrying out analyses of variance. Lindquist (**96**) and Lewis (**92**) discuss procedures for testing the goodness of fit of observed successive means to fitted curves in the case of a single group, but there is need for an expository article on the possibilities for comparing curvilinear trends for independent groups by variance methods.

Finally, it might be noted that the type of investigation in which all *S*s undergo several different experimental treatments in the same order should in general be avoided. The experiment by Bruner, Postman, and Mosteller (**13**) described in Table 9 is a case in point. The design is increased in sensitivity by the separation of individual variations in regression on time from the estimate of error, but this does not overcome the confounding of differences in "set" means with possible temporal effects of fatigue, adaptation, etc. Such confounding of the main experimental factor could have been obviated by randomizing the sequence of experimental conditions among the *S*s.

The Latin-Square Principle of Design

The fundamental principles of the latin-square design are described in many texts and in articles by Thomson (**127**), Grant (**52**), and Edwards (**38**). The latin square is essentially a triple classification, one variable represented by rows, the second by columns, and the third by treatments which occur once in each row and once in each column. As used by psychologists, the latin-square arrangement has typically been applied as a form of repeated-measurements design where *S*s are exposed to several experimental treatments and the investigator desires to take account of the possibility of systematic temporal effects such as transfer, practice, or fatigue.

Single latin-square designs. In the most common design using a single square, the criteria of classification are *S*s (rows), trials or successive periods (columns), and experimental treatments (Latin letters). While the latin square is a very compact form of

TABLE 10

NUMBER OF CORRECT RESPONSES MADE BY FIVE SUBJECTS IN READING THE LUCKIESH-MOSS LOW CONTRAST TEST CHART UNDER VARIOUS EXPERIMENTAL CONDITIONS (18)

	*Experimental Conditions**					
Subjects	*Loud Sound*	*Weak Sound*	*Heavy Pressure*	*Light Pressure*	*Control*	*Means*
A	21(2)	22(3)	20(4)	22(5)	22(1)	21.4
B	22(4)	16(1)	23(5)	19(2)	23(3)	20.6
C	14(1)	14(5)	23(2)	24(3)	20(4)	19.0
D	29(5)	24(4)	24(3)	24(1)	28(2)	25.8
E	16(3)	15(2)	14(1)	15(4)	13(5)	14.6
Mean	20.4	18.2	20.8	20.8	21.2	

	Experimental Days				
	1	2	3	4	5
Mean	18.0	21.2	21.8	20.2	20.2

Analysis of Variance

Source of Variation	*df*
Subjects	4
Days	4
Conditions	4
Residual (error)	12
Total	24

* The entries in parentheses are the days on which the experimental conditions were presented.

design, it should be obvious that its application assumes that interactions among the variables are negligible. If such interactions are present, they are confounded with the other sources of variation and may serve to augment or depreciate the apparent significances of effects. Chapanis, Rouse, and Schachter (**18**) employed a single 5×5 latin-square design in studying the effects of intersensory stimulation upon contrast sensitivity, as measured by number of correct responses on the Luckiesh-Moss Low Contrast Test Chart. Five *S*s were tested under five conditions (loud sound, weak sound, heavy pressure, light pressure, control) on each of five days. The analysis of results followed the form of Table 10.

Leyzorek (**91**) employed a single 7×7 latin-square design in a study analyzing various types of error scores made in visual interpolation between circular scale markers with differing sizes of scale interval.

Replicated latin-square designs. Studies in which latin-square designs are used may be replicated by employing randomly selected squares of the same size or by applying the same square to several groups of *S*s.

In some cases where the number of experimental treatments is small it may be feasible to apply all permutations of order of the several treatments with several *S*s undergoing each sequence. The principles involved in these variations have long been used in experimental research and special types of such designs have been variously referred to as permuted double-fatigue orders, balanced orders of presentation, rotation experiments, crossover designs, switchback studies, ABBA orders, etc. The relative advantages and disadvantages of various types of replication of the latin square and methods of analysis are discussed by Grant (**52**), Edwards (**37**, **38**), Cochran and Cox (**27**), and Kempthorne (**83**).

The simplest replicated latin-square design is the 2×2 square in which half of a group of *S*s go through two conditions in one order while the remaining *S*s go through the conditions in the reversed order (**53**). Frequently one of the conditions serves as an experimental control. Brogden (**12**), in a study of sensory conditioning, obtained auditory thresholds from 10 *S*s first in the presence of a light stimulus and then in the absence of light, while a second group of 10 *S*s was measured in the reversed sequence. Threshold measures were analyzed as in Table 11.

Similar designs were applied by Chernikoff and Brogden (**21**) and Chernikoff, Gregg, and Brogden (**22**) in studies of reaction time. In another study employing a 2×2 design of the same kind for a comparison of recall and recognition, the investigators inappropriately interpret the design as a 2×2 factorial (**113**).

Replication of the same latin-square design for larger squares follows a similar pattern of analysis with the addition of another source of variation entitled "square uniqueness" by Grant (**52**) or "latin-square

TABLE 11

AUDITORY THRESHOLDS WITH AND WITHOUT LIGHT (**12**)

	Experimental Group	
*Subject**	*Threshold with Light*	*Threshold without Light*
1	18.0	20.5
3	20.5	18.0
.	.	.
.	.	.
.	.	.
17	−4.5	5.5
19	15.0	15.0
Subgroup Mean	16.5	18.8
2	20.5	25.5
4	25.5	25.5
.	.	.
.	.	.
.	.	.
18	10.5	10.5
20	10.5	15.5
Subgroup Mean	15.5	18.2
Group Mean	16.0	18.5

Analysis of Variance

Source of Variation	*df*
Treatment	1
Ordinal position of treatment	1
Sequence of treatment	1
Individual variation of *S*s within sequences	18
Error	18
Total	39

* The odd-numbered *S*s make up the subgroup for which the threshold with light was made first and the even-numbered *S*s are the subgroup for whom the threshold with light was second.

TABLE 12

ERROR SCORES IN LINEAR PURSUIT AS A FUNCTION OF ANGLE OF ARM FROM BODY* (29)

Sequence	*Subject*	*Order*						
		I	II	III	IV	V	VI	VII
	1	(180°)	(210°)	(240°)	(270°)	(300°)	(330°)	(360°)
A	2	—	—	—	—	—	—	—
	3	—	—	—	—	—	—	—
	4	—	—	—	—	—	—	—
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
	25	(210°)	(240°)	(270°)	(300°)	(330°)	(360°)	(180°)
G	26	—	—	—	—	—	—	—
	27	—	—	—	—	—	—	—
	28	—	—	—	—	—	—	—

Analysis of Variance

Source of Variation	*df*
Angle	6
Sequence of angles	6
Ordinal position of angles	6
Individual differences within sequences	21
Square uniqueness	30
Remainder	126
Total	195

* Data in body of table not provided; entries indicate sequence of angles; each score was the mean of 20 trials at a given angle.

error" by Edwards (**38**). The latter author emphasizes application of tests of homogeneity of variance before pooling terms for an over-all error estimate. A 7×7 replicated latin-square design employing the same square for several groups of *S*s was used by Corrigan and Brogden (**29**) in studying the effect of bodily angle upon precision of linear pursuit movements. Twenty-eight *S*s were randomly allocated to seven groups of four *S*s each. Each group went through the pursuit task at seven angles with varying orders of presentation in a latin-square design. Error scores, based on combining 20 trials at each angle, were analyzed as in Table 12.

A similar analysis of an experimental design involving replication of the same 24×24 latin square with two *S*s in each sequence was carried out by Corrigan and Brogden (**30**). Two studies by Gregg and Brogden (**66, 67**) also involved the application of replicated 6×6 latin squares.

As indicated above, when the number of experimental treatments is small it is possible to utilize the latin-square principle by providing for every possible permutation of

order of presentation. This arrangement and the method of analysis are discussed by Grant (**52**). Ryan, Cottrell, and Bitterman (**118**) employed such a design in a study of muscular tension when they assigned four *S*s to each of the six possible orders of three experimental conditions (glare, noise, control), but their analysis of results was limited to treating the design as a double classification without replication. In an extension of the device of using every permutation of experimental orders, Grant, Jones, and Tallantis (**62**), studying concept formation by means of a card-sorting experiment, employed what might be called a double latin-square design since each group of *S*s repeated their assigned order of three experimental treatments. Under the conditions set up by the investigators, there were four *S*s for each of the 24 possible permutations specified.

The other major replicated latin-square design involves the application of several different randomly selected squares. No examples of this design other than those described by Grant (**52**) and Edwards (**38**) were found in the literature surveyed.

Combined latin-square and factorial design. The latin-square principle may be combined with a factorial arrangement of treatments in various ways (**37, 52, 27**). One simple method, for example, is to have a 4×4 square, with rows representing *S*s and columns representing successive trials, in which the four latin treatments are A_1B_1, A_1B_2, A_2B_1, and A_2B_2. This specific design was used by Prentice (**115**) in a study of the relation of distance to apparent size of figural after-effects. Four *S*s were subjected to the four different treatment combinations on four successive days in a latin-square design. Each treatment was the combination of one of two distances of stimulus and one of two conditions of "satiation." Analysis was carried out for only three sources of variation, viz., one *df* for the two variations of distance, seven (?) for individuals, and seven for error, but might have been extended as shown in Table 13.

TABLE 13

POINTS OF SUBJECTIVE EQUALITY FOR EACH SUBJECT FOR DIFFERENT DISTANCES AND CONDITIONS OF SATIATION* (**115**)

Subject	*Day* 1	2	3	4
A	2S	6NS	2NS	6S
B	2NS	6S	2S	6NS
C	6S	2S	6NS	2NS
D	6NS	2NS	6S	2S

Analysis of Variance

Source of Variation	*df*
Days	3
Subjects (sequences)	3
Satiation vs. no satiation	1
Distance	1
Interaction: Satiation × Distance	1
Remainder	6
Total	15

* Data not provided. The numbers in the body of table refer to distance (2 m. or 6 m.) at which *S* made his judgment; the letters S and NS refer to the conditions of "satiation" and "no satiation."

Another variation of the principle of combining latin-square and factorial design was found in a study of delayed response performance by Meyer, Harlow, and Settlage (**104**). In this experiment measures from sets of four monkeys were arranged in a 4×4 latin square, with four learning periods (rows) and four experimental conditions (columns). Within each of the 16 cells of the square there were 16 entries for all

combinations of four types of object pairs and four lengths of delay. Separate analyses were carried out for normal, unilateral damaged, and frontal damaged *S*s.

A third variation of combining latin-square and factorial principles in a replicated design was used by Cameron and Magaret (**16**) in a study of responses to incomplete sentences. In this study a 2×4 design was employed, each subcell indicating the combination of two factors. Replication of sequences by having seven *S*s undergo each order of conditions allowed for the differentiation of sequence variation from variation among *S*s in the same sequence.

A still more complex design involving latin-square and factorial principles in replication was used by Postman and Bruner (**114**) in a study of the relation of set and perceptual behavior. Their analysis, however, is difficult to explicate since composite sources of variation were dubiously partitioned and an attempt was made to analyze confounded interactions.

Greco-latin square designs. A further extension of the latin-square principle is to add another experimental treatment to the latin-square design in such a way that each new treatment appears but once with each Latin letter treatment (**44, 45, 52**). No "pure" examples of greco-latin square designs were found in the literature surveyed.

Comment. In a recent article McNemar (**102**) has discussed a type of application of the latin-square design which has been neglected by psychologists. As noted above, the common use of the latin square has been the case where one classification of the data consists of experimental treatments while the other two classifications consist of uncontrollable sources of variation, e.g., *S*s and trials. McNemar suggests the use of latin square as an economical form of three-factor design, when each factor consists of the same number of levels, and the mixed design where two of the three classifications are experimental factors. An example of the latter case is provided by Garrett and Zubin (**49**) who describe a study of color recognition by the dark-adapted eye where the three classifications in a 4×4 latin square were order of presentation (rows), levels of illumination (columns), and color (Latin letters). If the rows in this study had represented, say, four levels of dark adaptation, instead of order of presentation, this study could have served as an example of a three-factor latin-square design.

Of perhaps more importance is McNemar's contention that the latin-square design is rarely applicable in psychological research because the basic assumption of negligible interactions among the three classification variables is generally violated, especially in the design where *S*s form one of the criteria of classification. McNemar concludes that the use of the latin square is "defensible only in those rare instances when one has sound a priori reasons for believing that the interactions are zero" (**102**, p. 400).

There is no question but that the standard mathematical model of the latin-square design assumes that interactions are negligible and that statistical inference is most dependable when this assumption holds. But the writer does not agree with McNemar when he states that too many "significant" *F*'s are obtained when this assumption is not met because the residual term, containing both interaction and the ordinary error, tends to be smaller than the interaction properly used in the denominator for *F*. In the first place, it is clear that

the single latin square never provides an estimate of "pure" error of the type available when replication within the same subclasses is carried out. The residual of the latin square is always an admixture of confounded first-order and second-order interactions. This admixture, provided that the interactions are negligible, furnishes an unbiased estimate of "pure" error. When "significant" (but untestable within the design) interactions are present the residual could possibly be reduced in a specific experiment if the interaction(s) happened to follow the pattern of one or more of the three major classifications, but in the long run, if squares were always randomly selected, it would be expected that significant interaction(s) would tend to increase the residual, thus inflating the estimate of error. In such a case, if the experimenter were interested in testing the significance of a main effect against an estimate of "pure" error (regardless of the presence of interactions), there would be no such estimate available, and his purpose could not be met. If, unknowingly, a main effect were tested against such an inflated estimate of "pure" error, the F ratio would tend to be too small.

Let us assume, with McNemar, however, that the experimenter is desirous of testing the significance of main effects over and above the presence of possibly significant interactions. This would be analogous to testing main effects against interaction in the two-factor replicated design. The mathematical model for the two-factor case is simple because only one interaction is present. The important point, however, about the two-factor design is that the observed interaction term is assumed to be composed of two additive components, interaction variance plus error variance, while an observed main effect is assumed to be composed of three additive components, main effect variance plus interaction variance plus error variance. Since the residual term in the latin square is made up of several confounded interactions, it is impossible to set up a simple "components-of-variance" model (see below) as for the two-factor design. Nevertheless, for practical purposes one can assume that the observed residual of the latin square is composed of two components: confounded interactions variance plus error variance. Each observed main effect would then be assumed to consist of three components: main effect variance plus confounded interactions variance plus error variance. The consequent F ratio of main effect mean square over residual mean square should then tend to give an unbiased test of the significance of main effect over and above the presence of significant interactions. McNemar's contention that the denominator of the F test should properly consist of interaction alone, i.e., separated from error variance, has, so far as the writer is aware, no precedent in analysis of variance methods. In any case, however, it is clear that the presence of significant interactions negates the application of the usual latin-square design. No mathematical justification is readily available for inference in the case where some of the interactions are "significant."

Designs Involving Analysis of Covariance

In some investigations designed for analysis of variance it may not be feasible to control or classify the data on the basis of one or more relevant variables which can, however, be measured. The addition of covariance analysis (**44**) to the experimental design allows for adjustments to be made in experimental comparisons

on the basis of the regressions of the variable of primary importance on these other relevant variates. Covariance analysis may be carried out for all of the experimental designs so far presented. Discussions of covariance are readily available both for single-classification designs (**36**, **37**, **76**, **95**, **101**, **124**) and for multiple-classification designs (**37**, **76**, **95**, **124**), as well as for the case of one independent or control variable (**36**, **37**, **76**, **95**, **101**, **124**) or two independent variables (**76**, **124**). Snedecor (**124**) also provides an example of covariance in a latin-square design.

Applications of covariance analysis in experimental design were not too common in the literature surveyed. In general, moreover, when covariance was used little descriptive detail was provided. Bernberg (**9**) adjusted error scores for three groups of rats, learning a maze under three different conditions, on the basis of differential food intake in a single-classification design. In a study of reminiscence, Buxton and Bakan (**15**) adjusted criterion scores based on differences between rest and no-rest conditions by "correction" for recall trial difference scores. Buxton and Ross (**14**) similarly applied covariance analysis to a two-factor design in a study of the relationship between reminiscence and type of learning technique.[4] Reynolds (**117**), in a study of resistance to extinction, considered covariance adjustments of learning scores on the basis of scores on a previously trained habit, but rejected the plan because of heterogeneity of variances and low correlations. He then adjusted for training time in an analysis of the extinction scores. Glixman (**51**) applied covariance analysis to a $3\times3\times2$ factorial design in a study of recall of completed-incompleted tasks under differing conditions of stress. Covariance adjustments were made for scores based on number of incompleted-recalled tasks in terms of total number of incompleted tasks.

The complex problem of handling disproportionate subclass frequencies in a double-classification 2×4 factorial design with covariance analysis is exemplified in a study by Fitch, Drucker, and Norton (**46**), who used a procedure developed by Tsao (**128**). A rather full explanation of design, basic assumptions, and analytic procedures is provided by this study.

Comment. As noted by Fisher (**45**) and others, analysis of covariance in experimental design may be used for two major purposes: (*a*) to increase the precision of experimental comparisons by statistically controlling for sources of variation which do not lend themselves to experimental control, and (*b*) to aid in the interpretation of the results of an experiment. In the former case the experimenter should be sure that the "control" variable is independent of treatment effects (**45**). Ordinarily he is not particularly interested in studying the relationship between the concomitant measures and the primary variable being investigated (**95**). The standard example of a supplementary variable which can frequently be employed to improve the precision of an experiment is pretest scores on the same kind of performance which is to be measured during the experiment itself.

More care must be taken in deciding to utilize analysis of covariance for the second purpose. In this case supplementary measures are ordi-

[4] The main rationale for using covariance in this study was to "remove" variance due to using the same *Ss* under experimental and control conditions. Grant (**52**) illustrates how this same study might have been analyzed by interpreting the arrangement as a 2×2 greco-latin square.

narily taken during the course of the experiment and hence variations in the concomitant variable may be a function of the experimental treatments. The obvious difficulty in applying analysis of covariance here is that adjustment of the primary variable may remove part of the treatment effect itself. The experimenter may wish, however, to find out whether there are significant differences in treatment effects on the primary variable when the secondary variable is "equalized" over all groups. In such cases it is generally profitable to carry out not only an analysis of covariance to "eliminate" possible effects of the secondary variable, but also separate analyses of variance of both the secondary variable and the unadjusted primary measures, as well as careful examination of regression and correlation coefficients. Comparison of the several analyses will tend to clarify the extent to which experimental effects on the primary variable act directly or indirectly through the mediation of the concomitant variable or covariate. For example, to paraphrase Snedecor (**124**, p. 335), in the study by Glixman (**51**) cited above: Did the *S*s show differences in number of incompleted-recalled tasks under varying conditions of stress because of differences in total number of incompleted tasks, or in spite of them? Excellent discussions of the use of analysis of covariance to improve understanding of experimental structure are presented by Edwards (**37**) and others (**27, 83, 124**).

In addition to the complexities of procedure and interpretation which generally arise when analysis of covariance is applied to designs involving multiple classification, or when there is more than one concomitant variable, other difficulties sometimes arise in the use of covariance analysis. At times the regression of primary variable upon supplementary variable may be nonlinear, necessitating the adjustment of primary variable on the basis of curvilinear regression (**80**). In other cases the experimenter may discover that there are problems in the choice of appropriate regression coefficients for estimation of the main variable because of marked heterogeneity of regression from subclass to subclass. Jackson (**73, 74**) provides a detailed discussion of such problems and others and suggests possible solutions.

In the context of analysis of covariance, special mention should be made of the Johnson-Neyman technique (**78, 81**). As noted above, one of the major uses of covariance methods is to adjust experimental comparisons for extraneous causes of variation. Frequently, such adjustment is designed to "equate" experimental groups when it is not feasible to increase precision by pairing cases or otherwise matching the several groups with respect to the relevant measures. The Johnson-Neyman technique not only furnishes a test of whether a statistically significant difference exists between the means of the groups being compared but, in addition, specifies the range of control variables for which a conclusion of significant difference may be regarded to hold. Moreover, no special difficulties are involved when the groups are unequal in number. Johnson and Fay (**81**) provide the detailed computational and graphical solution for a problem in which the social studies achievement of 90 pupils who excel in the ability to predict the outcome of given events is compared with the social studies achievement of 90 pupils who are poor predictors. The null hypothesis rejected on the basis of the analysis was that no difference exists in mean achievement

between superior and inferior predictors when the effects of chronological and mental ages are controlled. The unique surplus information contributed by the Johnson-Neyman technique indicated the range of mental age and chronological age for which the conclusion of significant difference was valid.

General Considerations

Models, assumptions, and transformations in analysis of variance. Before valid inferences may be drawn from an analysis of variance, the data must reasonably satisfy certain assumptions made about the underlying mathematical models used in the analysis and subsequent tests of significance. In recent years various sets of assumptions have been proposed about the elements in the linear models whereby analysis of variance is used for statistical inference. Pointing out that Fisher (**44**) had originally introduced the twofold conception, Eisenhart (**41**), in 1947, elaborated on the viewpoint that analysis of variance involves one of two basic models, each appropriate for the solution of a different class of problems: *Model I* to detect or estimate fixed relations among population means, and *Model II* to detect or estimate components of random variation ascribable to the different factors being investigated. The former is frequently referred to as the ***standard*** model while the latter is commonly called the ***components-of-variance*** model (**105**).

In brief the major distinction between the two models is that Model I assumes that treatment and other designated effects are additive fixed constants, introducing systematic variation, while Model II assumes that treatment and other effects are random variables each having a normal distribution. Both models assume that experimental (residual) errors are independently and normally distributed with a constant variance. The decision as to whether a given element in the linear model is best represented by a ***mean***, indicating a systematic source of variation, or by a ***variance***, indicating a random source of variation, depends upon the extent to which the respective variable was randomly sampled. In many experiments some of the effects are best regarded as fixed, e.g., the usual case for experimental treatments which are rarely randomly drawn from a population of possible treatments, while other effects may be regarded as introducing random variation, e.g., effects assignable to *S*s drawn at random from a specified population. When both types of elements are present, the underlying model is described as "mixed."[6]

The majority of published psychological studies employing variance designs give little evidence that investigators pay much attention to the several assumptions underlying analysis of variance. For instance, although tests of homogeneity of subgroup variance are readily available, e.g., Bartlett's test (**37, 76, 84, 124**), the L_1 test (**76, 84**), the *M* test (**76, 84**), and Box's test (**36**), the assumption that experimental errors have

[6] A recent review by Crump (**31**) indicates that Eisenhart's so-called Model I, Model II, and Mixed Model have been supplemented by Tukey with Models III, IV, V, and X, all involving somewhat different assumptions. Many of the analyses described by Kempthorne (**83**) are based upon finite "randomization" models, involving no assumption about normality of error distributions. Other models have been proposed which are also nonparametric, i.e., make no assumption about the form of the population distributions (**105, 106**). In general, the assumptions involved in all of these models are less restrictive than the usual set of assumptions, thus broadening the potential applicability of analysis of variance techniques.

equal variance appears to be tested only in a minority of experiments. The assumptions that errors are uncorrelated, with constant variance, appear on both empirical and theoretical grounds to be somewhat more critical than the assumption of normality of distribution of the errors (**25**, **27**, **83**).

Cochran, in a detailed discussion of the consequences when the assumptions for analysis of variance as a technique for carrying out tests of significance of differences among means are not satisfied, states that "the principal methods for an improved analysis are omission of certain observations, treatments, or replicates, subdivision of the error variance, and transformation to another scale before analysis" (**25**, p. 37). The method most frequently resorted to by psychologists who take cognizance of violation of assumptions is transformation of the scale. The rationale and conditions for various kinds of transformations are most fully discussed by Bartlett (**5**). Briefer accounts are available in other sources (**37**, **76**, **83**, **124**). Such transformations are frequently intended to stabilize error variance, especially in cases where variances within the subclasses show a functional relationship with subclass means.

In the experimental literature, Haggard (**68**, **69**) provides a careful investigation of the problem of selecting proper measures of GSR data for analysis of variance procedures and the effects of using inappropriate measures. Among the specific transformations that were utilized in recent psychological studies were square-root transformation of number of reversals of perspective (**13**), log transformation of latency scores (**4**, **93**), log transformation of hoarding scores (**72**), log transformation of number of contacts in a pursuitmeter task (**30**), reciprocal transformation of latency scores (**89**), arcsine transformation of percentages (**75**, **94**, **109**), and transformation of obtained scores to per cent of prestimulus values (**35**).

The F test in analysis of variance. In the single-classification design the error term of the *F* ratio is provided by the "within-groups" mean square. Similarly for the double-classification design with a single observation in each cell, the denominator of the *F* ratio is furnished by the interaction or remainder mean square. Complications arise, however, in selecting the proper denominators for the *F* ratios in the case of the double-classification design with several observations in each subclass. In general, the most widely used procedure has depended upon whether or not the investigator decides that the interaction *F* test is "significant." If the interaction *F* test is not found to be significant, the investigator may either use the "within-cells" mean square as his error estimate in testing the significance of variation among main effects or pool the "interaction" and "within-cells" *SS* in arriving at an error estimate. If, however, he finds that the interaction mean square is significant, he generally employs the latter square in his *F* tests for the main criteria of classification. Strictly speaking, the use of the interaction mean square as an estimate of error in this case involves application of a "components-of-variance" model to the data, since it is thereby assumed that the expected value of the mean square for an apparent main effect is made up of a linear combination of error variance, interaction variance, and the intrinsic main effect variance itself. The hypothesis is thus being tested that the intrinsic main effect is not

significant over and above any variation attributable to both random error and the effect of interaction.

With higher order multiple-classification or factorial designs, the selection of appropriate F tests becomes more complex. Again, psychologists have generally proceeded "from the bottom up" in setting up F ratios from an analysis of variance table. For example, if a three-factor design with several replications per subclass is involved, the first test is generally made by setting up the F ratio of the highest order interaction mean square over the "within-cells" mean square. If this highest order interaction is found to be "not significant," it may then be pooled with the "within-cells" term to provide the denominator for F tests of the next highest interaction terms. On the other hand, if the highest order interaction is found to be "significant," the F tests of the next highest interaction mean squares are made with the highest order interaction mean square in the denominators. When the investigator arrives at the main variables of classification he generally has used as his denominator for F ratios the pooled interactions and residual, if none of the preceding F's has been significant, or the interaction mean square of largest magnitude which contains the elements assumed to be contributing to the apparent variability of a given main variable. This *ad hoc*, somewhat intuitive procedure for arriving at F ratios may frequently be criticized from the standpoint of the variance components assumed to be operating in the specific situation or because the elements assumed to be random variables in the linear model may more logically be assumed to represent fixed parameters. If, however, a "components-of-variance" model (**41**) is justified by the data and sampling methods used in the study, a more appropriate method of testing relations is available. Among others (**31**, **83**, **84**, **105**), Cochran (**26**) has recently discussed in detail the problems arising in testing a null hypothesis about several means when an appropriate denominator for an F ratio is not immediately provided by the "expected mean squares" in the analysis of variance table. The procedure suggested involves setting up what Cochran calls an F' test where numerator and denominator of the F ratio are linear combinations of mean-square terms arranged in such a way that the treatment effect to be tested is present only in the numerator, while all remaining assumed components of variance are present in both the numerator and denominator. The respective df's for the composite ratio are determined according to approximations proposed by Satterthwaite (**119**, **120**). The formulation of such ratios is facilitated by the provision of expected mean squares for many commonly used experimental designs by Snedecor (**124**) and Mood (**105**).

Individual tests of significance in the analysis of variance. Investigators frequently desire to follow an over-all analysis of variance with tests of the significances of differences between individual pairs or groups of treatment means. When the F associated with a given classification is found to be significant, the most commonly used procedure has been the method described by Lindquist (**95**), in which t tests are applied to the selected means, using standard errors of differences based on the appropriate error variance from the analysis of variance. Confidence intervals can be set up on the same basis. When the F ratio representing a given classification is found to be not significant, the investigator generally ceases his anal-

ysis. In recent years this simple alternative operation has been criticized and further extensions of analysis of variance have been suggested. Dixon and Massey (**36**) propose a "test for extreme mean" applicable in the situation where one group of *S*s is a control group, while the remaining groups are experimental groups. Johnson (**76**), utilizing a suggestion made by Fisher (**45**), discusses how *selected* pairs of means may be compared by lowering the p level for significance in accordance with the possible number of comparisons. Snedecor (**124**) and Cochran and Cox (**27**) warn about the dangers of testing differences suggested by the data and present methods for subdividing the treatment and error *SS* for relevant individual and group comparisons.

Subdivision of treatment *SS* is especially applicable in experiments where the levels of a given factor represent varying amounts or categories along a treatment continuum, e.g., degrees of learning. Kelman (**82**), for example, had four groups of *S*s (Control, Success, Failure, Ambiguous) which furnished three orthogonal comparisons: (*a*) Control group vs. experimental groups (3C-S-F-A); (*b*) Ambiguous reinforcement vs. clear-cut Success or Failure (2A-S-F); and (*c*) Success vs. Failure (S-F). Johnson and Tsao (**79**), in a $4\times7\times2\times2\times2$ factorial study dealing with the determination of differential limen values, furnish a detailed discussion of the application of orthogonal polynomials (**44**) in expressing the relationships between the factors, e.g., weight, rate, etc., and the limen values. The procedure of fitting orthogonal polynomials to a given factorial classification with associated tests of significance for linear regression, parabolic regression, etc. can frequently be used to furnish information and answers to questions which are not supplied by the over-all F test for a given set of treatment means (**27**, **124**).

Perhaps the most simple and practical procedure for comparing individual means now available to the investigator who is not satisfied with the results of an over-all F test is that presented by Tukey (**129**). In this procedure, after finding a significant F for a set of treatment means, one successively applies a "gap" test, a "straggler" test, and a new F test to subgroups among the treatment means to detect distinguishable groups.

The power of analysis of variance tests. Almost universally, the psychologist has limited his attention in testing hypotheses to consideration of errors *of the first kind* (Type I errors), i.e., rejecting hypotheses when they are true. The risk of committing errors *of the second kind* (Type II errors), i.e., accepting false hypotheses, has in general entered very little into his schema of statistical inference. The usual test of significance involves the specification of a so-called critical region which controls only the risk of error of the first kind. Thus, for example, if the critical region is set at the .05 level for an F test, the experimenter is in effect declaring that rejection of the null hypothesis when p equals or exceeds this value will be wrong only 5 per cent of the time.

What happens, however, when a null hypothesis is not rejected, i.e., if the F ratio is smaller than, say, the tabulated value for the .05 level? Most investigators appreciate in theory that such a finding does not mean that the null hypothesis is proved. And yet the tendency is strong to accept the null hypothesis and draw the conclusion that no differences among means are present. The *power function* of a given test of significance is

designed to indicate the probabilities of rejecting a specified hypothesis when alternative hypotheses are assumed to be true. In the usual case where the specified hypothesis is a null hypothesis, i.e., all of a group of means are equal, the probability of rejecting the null hypothesis when the true means are in fact different depends upon the significance level selected for the test, the magnitude of the differences among the means, the size of the error variance, and the number of replicates.

Two major approaches have been devised for determining the power of analysis of variance tests. In the older method developed by Tang (**126**), the alternative hypothesis to the null hypothesis is expressed in terms of the variance of a finite set of assumed population means equal in number to the number of observed means involved in the *F* test. In a more recent approach described by Ferris, Grubbs, and Weaver (**43**), the alternative hypothesis is expressed in terms of a set of normally distributed population means, these means representing a sample from a normal superpopulation with variance bearing a specified ratio to the error variance involved in the *F* test. Since, however, this latter paper presents a somewhat limited set of curves for estimating the power of the analysis of variance test at only the .05 significance level, further discussion will be limited to the Tang approach.[6]

The method developed by Tang assumes that the observations can be expressed in terms of Model I (see above), which assumes a linear combination of mean effects and errors which are normally and independently distributed with constant variance. Tang presents fairly extensive tables for varying pairs of *df*, under the assumption that either a .05 or .01 level of significance is being employed for rejection of the null hypothesis. In these tables the probabilities of error of the second kind are indicated for varying sizes of ϕ, a variance ratio with numerator derived from the assumed alternative hypothesis. Lehmer (**90**) subsequently prepared tables providing the value of ϕ required for a specified probability of error of the second kind. Tang's tables are reproduced with extensive discussion of their use in Kempthorne (**83**) and Mann (**99**), while Lehmer's tables for Type II errors of probability .3 and .2 are available in Dixon and Massey (**36**).

The power function of analysis of variance tests is very useful (*a*) for estimating the sample sizes that will reasonably guarantee a desired probability of error of the second kind for a specified alternative hypothesis and a designated level of significance, and (*b*) for determining the power of a test, if the sample sizes and significance level have already been fixed. Perhaps the chief implication of the power concept in relation to current psychological research is the conclusion that many experiments are carried out without sufficient replication to insure a reasonable chance of detecting experimentally important differences in treatment means. Kempthorne (**83**), for example, estimates that six replicates in each subclass are necessary in a $2 \times 2 \times 2$ factorial experiment to insure with a probability of .95 that a true difference of the two means for a given factor equal in size to the error standard deviation will

[6] Actually the paper of Ferris, Grubbs, and Weaver (**43**) presents "operating-characteristics" curves which are related to power curves as x is to $1-x$, i.e., complementary. Eisenhart *et al.* (**42**) also provide a brief discussion of operating-characteristic functions for analysis of variance tests based upon Eisenhart's Model II.

be detected (using a .05 test of significance).

The power conception is somewhat contrary to the commonly accepted notion that an experimenter should insist on a very stringent level of significance before rejecting a null hypothesis when his samples are relatively small. Such a notion, paradoxically defended on the grounds of conservatism, has probably resulted in the premature dismissal of many potentially important areas of experimentation. It might be a worthwhile addendum to current methodology if many exploratory, small-sample experiments were primarily devoted, not to tests of hypotheses, but to obtaining estimates of error variance. Once such estimates of error variance have been obtained, the experimenter is in a position to determine the sample sizes necessary to detect differences between treatments regarded to be of practical or theoretical significance.

An overview of psychological research design and analysis. Because of the widely accepted thesis that experimental design and statistical analysis are "dynamic" aspects of the same research "whole," and inasmuch as many statistical and measurement techniques, e.g., regression and correlational analysis, the t test, chi square, discriminant functions, etc., can be subsumed under analysis of variance and the F distribution, a thorough survey of variance designs used by psychologists could have been extended far beyond the limits of the present article. Minimum attention, moreover, has been given to the general theory of experimental design in this survey. Basic concepts such as experimental control, statistical control, randomization, replication, balance, efficiency, precision, orthogonality, comprehensiveness, self-containedness, etc. enter in the adoption of any specific experimental design, but, in general, the journal reporting of studies is not amenable to elaboration of underlying principles.

With regard to specific experimental arrangement, cursory examination of books which have treated experimental design in agricultural, biological, or industrial research, from Fisher's classic (**45**) to the recent comprehensive presentations by Cochran and Cox (**27**) and Kempthorne (**83**), reveals that psychologists have generally utilized only the simpler and "complete" experimental configurations. Of the 150 experimental plans presented by Cochran and Cox (**27**), only a small minority seem to have appeared in psychological research design. As noted previously, various devices of ***deliberate confounding*** or ***partial confounding***, especially applicable in higher-order factorial studies, seem to be only rarely considered, despite their early introduction into the psychological literature by Baxter (**6**) and their potential experimental and practical advantages. This is not to say that methodology from one area of research can be routinely applied in another area, but the increasing frequency with which variance designs have been applied in psychological research probably indicates a trend which can be expected to continue.

The practicing researcher, of course, finds difficulties in keeping up with current developments and refinements in the area of experimental design and analysis. A real service is being performed by the excellent summaries being presented in the *Annual Review of Psychology* (**39, 54, 103**). Nevertheless, as pointed out by Johnson (**77**) in a recent discussion of the contribution of statistical science to educational and psychological re-

search, the "newer developments in the field have mainly been specialized devices for specialized purposes" with three basic principles of experimentation—replication, randomization, and control of variability—being the foundation stones of modern experimental design. Current psychological research, as we have seen, has been tremendously influenced by the "Fisherian revolution in methods of experimentation" (**133**).

Summary

This article has presented a survey of the major types of experimental design involving analysis of variance which have characterized psychological research during recent years. The survey is implemented by brief reference to specific studies utilizing a variety of experimental configurations which have appeared in the literature. Some comments were made about the appropriateness of design or analysis in particular instances, followed by a discussion of general considerations in application of variance design and analysis.

BIBLIOGRAPHY

1. Alexander, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.
2. Ammons, R. B. Effects of pre-practice activities on rotary pursuit performance. *J. exp. Psychol.*, 1951, **41**, 187–191.
3. Anderson, R. L. Missing-plot techniques. *Biometrics*, 1946, **2**, 41–47.
4. Arnold, W. J. Simple reaction chains and their integration. II. Heterogeneous chaining with terminal reinforcement. *J. comp. physiol. Psychol.*, 1947, **40**, 427–440.
5. Bartlett, M. S. The use of transformations. *Biometrics*, 1947, **3**, 39–52.
6. Baxter, B. Problems in the planning of psychological experiments. *Amer. J. Psychol.*, 1941, **54**, 270–280.
7. Beebe-Center, J. G., Black, P., Hoffman, A. C., & Wade, M. Relative per diem consumption as a measure of preference in the rat. *J. comp. physiol. Psychol.*, 1948, **41**, 239–251.
8. Bendig, A. W., & Braun, H. W. The influence of the genotype on the retention of maze habit in the rat following electroshock convulsions. *J. comp. physiol. Psychol.*, 1951, **44**, 112–117.
9. Bernberg, R. E. A comparison of the effects of electroconvulsive shock and electronarcosis upon the learning ability of young rats. *J. comp. physiol. Psychol.*, 1951, **44**, 50–60.
10. Black, J. W. A compensatory effect in vocal responses to stimuli of low intensity. *J. exp. Psychol.*, 1950, **40**, 396–397.
11. Bray, D. W. The prediction of behavior from two attitude scales. *J. abnorm. soc. Psychol.*, 1950, **45**, 64–84.
12. Brogden, W. J. Sensory conditioning measured by the facilitation of auditory acuity. *J. exp. Psychol.*, 1950, **40**, 512–519.
13. Bruner, J. S., Postman, L., & Mosteller, F. A note on the measurement of reversals of perspective. *Psychometrika*, 1950, **15**, 63–72.
14. Buxton, C. E., & Ross, H. V. Relationship between reminiscence and type of learning technique in serial anticipation learning. *J. exp. Psychol.*, 1949, **39**, 41–46.
15. Buxton, C. E., & Bakan, M. B. Correction vs. non-correction learning techniques as related to reminiscence in serial anticipation learning. *J. exp. Psychol.*, 1949, **39**, 338–341.
16. Cameron, N., & Magaret, Ann. Experimental studies in thinking: I. Scattered speech in the responses of normal subjects to incomplete sentences. *J. exp. Psychol.*, 1949, **39**, 617–627.
17. Carpenter, A. The rate of blinking during prolonged visual search. *J. exp. Psychol.*, 1948, **38**, 587–591.
18. Chapanis, A., Rouse, R. O., & Schachter, S. The effect of inter-sensory stimulation on dark adaptation and night vision. *J. exp. Psychol.*, 1949, **39**, 425–437.
19. Chapanis, A., & Leyzorek, M. Accuracy of visual interpolation between scale markers as a function of the num-

ber assigned to the scale. *J. exp. Psychol.*, 1950, **40**, 655–667.

20. CHERNIKOFF, R., & BROGDEN, W. J. The effect of instructions upon sensory pre-conditioning of human subjects. *J. exp. Psychol.*, 1949, **39**, 200–207.
21. CHERNIKOFF, R., & BROGDEN, W. J. The effect of response termination of the stimulus upon reaction time. *J. comp. physiol. Psychol.*, 1949, **42**, 357–364.
22. CHERNIKOFF, R., GREGG, L. W., & BROGDEN, W. J. The effect of fixed duration stimulus magnitude upon reaction time to a response terminated stimulus. *J. comp. physiol. Psychol.*, 1950, **43**, 123–128.
23. CHILD, I. L., & GROSSLIGHT, J. H. The effect of substitute activity as depending upon the nature of the similarity between substitute and original activity. *Amer. J. Psychol.*, 1947, **60**, 226–239.
24. CITRON, A. F., CHEIN, I., & HARDING, J. Anti-minority remarks: a problem for action research. *J. abnorm. soc. Psychol.*, 1950, **45**, 99–126.
25. COCHRAN, W. G. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, 1947, **3**, 22–38.
26. COCHRAN, W. G. Testing a linear relation among variances. *Biometrics*, 1951, **7**, 17–32.
27. COCHRAN, W. G., & COX, G. M. *Experimental designs.* New York: Wiley, 1950.
28. CONKLIN, J. E. Three factors affecting the general level of electrical skin resistance. *Amer. J. Psychol.*, 1951, **64**, 78–86.
29. CORRIGAN, R. E., & BROGDEN, W. J. The effect of angle upon precision of linear pursuit movements. *Amer. J. Psychol.*, 1948, **61**, 502–510.
30. CORRIGAN, R. E., & BROGDEN, W. J. The trigonometric relationship of precision and angle of linear pursuit movements. *Amer. J. Psychol.*, 1949, **62**, 90–98.
31. CRUMP, S. L. The present status of variance component analysis. *Biometrics*, 1951, **7**, 1–16.
32. CRUTCHFIELD, R. S. Efficient factorial design and analysis of variance illustrated in psychological experimentation. *J. Psychol.*, 1938, **5**, 339–346.
33. CRUTCHFIELD, R. S. The determiners of energy expenditure in string-pulling by the rat. *J. Psychol.*, 1939, **7**, 163–178.
34. CRUTCHFIELD, R. S., & TOLMAN, E. C. Multiple-variable design for experiments involving interaction of behavior. *Psychol. Rev.*, 1940, **47**, 38–42.
35. DAVIS, R. C. Motor responses to auditory stimuli above and below threshold. *J. exp. Psychol.*, 1950, **40**, 107–120.
36. DIXON, W. J., & MASSEY, F. J. *Introduction to statistical analysis.* New York: McGraw-Hill, 1951.
37. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
38. EDWARDS, A. L. Homogeneity of variance and the Latin square design. *Psychol. Bull.*, 1950, **47**, 118–129.
39. EDWARDS, A. L. Statistical theory and research design. In C. P. Stone (Ed.), *Annual review of psychology.* Stanford: Annual Reviews, 1951. Pp. 335–352.
40. EDWARDS, A. L., & HORST, P. The calculation of sums of squares for interactions in the analysis of variance. *Psychometrika*, 1950, **15**, 17–24.
41. EISENHART, C. The assumptions underlying the analysis of variance. *Biometrics*, 1947, **3**, 1–21.
42. EISENHART, C., HASTAY, M. W., & WALLIS, W. A. *Selected techniques of statistical analysis.* New York: McGraw-Hill, 1947.
43. FERRIS, C. D., GRUBBS, F. E., & WEAVER, C. L. Operating characteristics for the common statistical tests of significance. *Ann. math. Statist.*, 1946, **17**, 178–197.
44. FISHER, R. A. *Statistical methods for research workers.* Edinburgh: Oliver and Boyd, 1925–1946; New York: Hafner, 1950.
45. FISHER, R. A. *The design of experiments.* Edinburgh: Oliver and Boyd, 1935–1951.
46. FITCH, M. L., DRUCKER, A. J., & NORTON, J. A. Frequent testing as a motivating factor in large lecture classes. *J. educ. Psychol.*, 1951, **42**, 1–20.
47. FRANKLIN, J. D., & BROŽEK, J. The relation between distribution of practice and learning efficiency in psychomotor performance. *J. exp. Psychol.*, 1947, **37**, 16–24.
48. FURCHTGOTT, E. Effects of total body X-irradiation in learning: an exploratory study. *J. comp. physiol. Psychol.*, 1951, **44**, 197–203.
49. GARRETT, H. E., & ZUBIN, J. The analysis of variance in psychological research. *Psychol. Bull.*, 1943, **40**, 233–267.

50. Gebhard, M. E. The effect of success and failure upon the attractiveness of activities as a function of experience, expectation, and need. *J. exp. Psychol.*, 1948, **38**, 371–388.

51. Glixman, A. F. Recall of completed and incompleted activities under varying degrees of stress. *J. exp. Psychol.*, 1949, **39**, 281–295.

52. Grant, D. A. The Latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, **45**, 427–442.

53. Grant, D. A. The statistical analysis of a frequent experimental design. *Amer. J. Psychol.*, 1949, **62**, 119–122.

54. Grant, D. A. Statistical theory and research design. In C. P. Stone (Ed.), *Annual review of psychology*. Stanford: Annual Reviews, 1950. Pp. 277–296.

55. Grant, D. A. Perceptual versus analytical responses to the number concept of a Weigl-type card-sorting test. *J. exp. Psychol.*, 1951, **41**, 23–29.

56. Grant, D. A., & Norris, E. B. Eyelid conditioning as influenced by the presence of sensitized beta-responses. *J. exp. Psychol.*, 1947, **37**, 423–433.

57. Grant, D. A., Norris, E. B., & Boissard, S. Dark adaptation and the pseudo-conditioned eyelid response. *J. exp. Psychol.*, 1947, **37**, 434–439.

58. Grant, D. A., Hake, H. W., & Schneider, D. E. Effects of pre-testing with the conditioned stimulus upon extinction of the conditioned eyelid response. *Amer. J. Psychol.*, 1948, **51**, 243–246.

59. Grant, D. A., & Schneider, D. E. Intensity of the conditioned stimulus and strength of conditioning: I. The conditioned eyelid response to light. *J. exp. Psychol.*, 1948, **38**, 690–696.

60. Grant, D. A., & Schneider, D. E. Intensity of the conditioned stimulus and strength of conditioning: II. The conditioned galvanic skin response to an auditory stimulus. *J. exp. Psychol.*, 1949, **39**, 35–40.

61. Grant, D. A., Hornseth, J. P., & Hake, H. W. Sensitization of the beta-response as a function of the wave length of the stimulus. *J. exp. Psychol.*, 1949, **39**, 195–199.

62. Grant, D. A., Jones, O. R., & Tallantis, B. The relative difficulty of the number, form, and color concepts of a Weigl-type problem. *J. exp. Psychol.*, 1949, **39**, 552–557.

63. Grant, D. A., & Mote, F. A. Effects of brief flashes of light upon the course of dark-adaptation. *J. exp. Psychol.*, 1949, **39**, 610–616.

64. Grant, D. A., Riopelle, A. J., & Hake, H. W. Resistance to extinction and the pattern of reinforcement. I. Alternation of reinforcement and the conditioned eyelid response. *J. exp. Psychol.*, 1950, **40**, 53–60.

65. Grant, D. A., Hornseth, J. P., & Hake, H. W. The influence of the inter-trial interval on the Humphreys' "random reinforcement" effect during the extinction of a verbal response. *J. exp. Psychol.*, 1950, **40**, 609–612.

66. Gregg, L. W., & Brogden, W. J. The relation between duration and reaction time difference to fixed duration and response terminated stimuli. *J. comp. physiol. Psychol.*, 1950, **43**, 329–337.

67. Gregg, L. W., & Brogden, W. J. The relation between reaction time and the duration of the auditory stimulus. *J. comp. physiol. Psychol.*, 1950, **43**, 389–395.

68. Haggard, E. A. On the application of analysis of variance to GSR data: I. The selection of an appropriate measure. *J. exp. Psychol.*, 1949, **39**, 378–392.

69. Haggard, E. A. On the application of analysis of variance to GSR data: II. Some effects of the use of inappropriate measures. *J. exp. Psychol.*, 1949, **39**, 861–867.

70. Helson, H. Design of equipment and optimal human operation. *Amer. J. Psychol.*, 1949, **62**, 473–497.

71. Horowitz, M. W. An analysis of the superiority of binocular over monocular visual acuity. *J. exp. Psychol.*, 1949, **39**, 581–596.

72. Hunt, J. McV., Schlosberg, H., Solomon, R. L., & Stellar, E. Studies of the effects of infantile experience on adult behavior in rats. I. Effects of infantile feeding frustration on adult hoarding. *J. comp. physiol. Psychol.*, 1947, **40**, 291–304.

73. Jackson, R. W. B. Applications of the analysis of variance and covariance method to educational problems. *Dept. Educ. Res. Univer. Toronto Bull.*, 1940, No. 11.

74. Jackson, R. W. B. Some difficulties in the application of the analysis of covariance method to educational problems. *J. educ. Psychol.*, 1941, **32**, 414–422.

75. Jenkins, W. O., & Postman, L. An experimental analysis of set in rote learning: retroactive inhibition of

changing set. *J. exp. Psychol.*, 1949, **39**, 69–72.

76. JOHNSON, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.

77. JOHNSON, P. O. Modern statistical science and its function in educational and psychological research. *Sci. Mon.*, 1951, **62**, 385–396.

78. JOHNSON, P. O., & NEYMAN, J. Tests of certain linear hypotheses and their application to some educational problems. *Statist. res. Mem.*, 1933, **1**, 72–93.

79. JOHNSON, P. O., & TSAO, F. Factorial design in the determination of differential limen values. *Psychometrika*, 1944, **9**, 107–144.

80. JOHNSON, P. O., & TSAO, F. Factorial design and covariance in the study of individual educational development. *Psychometrika*, 1945, **10**, 133–162.

81. JOHNSON, P. O., & FAY, L. C. The Johnson-Neyman technique, its theory and application. *Psychometrika*, 1950, **15**, 349–367.

82. KELMAN, H. C. Effects of success and failure on "suggestibility" in the autokinetic situation. *J. abnorm. soc. Psychol.*, 1950, **45**, 267–285.

83. KEMPTHORNE, O. *The design and analysis of experiments.* New York: Wiley, 1952.

84. KENNEY, J. F., & KEEPING, E. S. *Mathematics of statistics.* Part Two. New York: Nostrand, 1951.

85. KIMBLE, G. A., & BILODEAU, E. A. Work and rest as variables in cyclical motor learning. *J. exp. Psychol.*, 1949, **39**, 150–157.

86. KOGAN, L. S. Analysis of variance-repeated measurements. *Psychol. Bull.*, 1948, **45**, 131–143.

87. KUNTZ, J. E., & SLEIGHT, R. B. Legibility of numerals: the optimum ratio of height to width of stroke. *Amer. J. Psychol.*, 1950, **63**, 567–575.

88. LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.

89. LAWRENCE, D. H., & MILLER, N. E. A positive relationship between reinforcement and resistance to extinction produced by removing a source of confusion for a technique that had produced opposite results. *J. exp. Psychol.*, 1947, **37**, 494–509.

90. LEHMER, E. Inverse tables of probabilities of errors of the second kind. *Ann. math. Statist.*, 1944, **15**, 388–398.

91. LEYZOREK, M. Accuracy of visua interpolation between circular scale markers as a function of the separation between markers. *J. exp. Psychol.*, 1949, **39**, 270–271.

92. LEWIS, D. *Quantitative methods in psychology.* Iowa City: The Bookshop, 1949.

93. LIBERMAN, A. M. A comparison of transfer effects during acquisition and extinction of two instrumental responses. *J. exp. Psychol.*, 1951, **41**, 192–198.

94. LICKLIDER, J. C. R., BINDRA, D., & POLLACK, I. The intelligibility of rectangular speech-waves. *Amer. J. Psychol.*, 1948, **61**, 1–20.

95. LINDQUIST, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.

96. LINDQUIST, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.

97. LITTMAN, R. A. Conditioned generalization of the galvanic skin reaction to tones. *J. exp. Psychol.*, 1949, **39**, 868–882.

98. LITTMAN, R. A. Latent learning in a T-maze after two degrees of training. *J. comp. physiol. Psychol.*, 1950, **43**, 135–147.

99. MANN, H. B. *Analysis and design of experiments.* New York: Dover, 1949.

100. MANN, C. W., & PASSEY, G. E. The perception of the vertical: V. Adjustment to the postural vertical as a function of the magnitude of postural tilt and duration of exposure. *J. exp. Psychol.*, 1951, **41**, 108–113.

101. MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1949.

102. MCNEMAR, Q. On the use of latin squares in psychology. *Psychol. Bull.*, 1951, **48**, 398–401.

103. MCNEMAR, Q. Statistical theory and research design. In C. P. Stone (Ed.), *Annual review of psychology.* Stanford: Annual Reviews, 1952. Pp. 409–418.

104. MEYER, D. R., HARLOW, H. F., & SETTLAGE, P. H. A survey of delayed response performance by normal and brain damaged monkeys. *J. comp. physiol. Psychol.*, 1951, **44**, 17–25.

105. MOOD, A. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.

106. MOSES, L. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122–143.

107. NEWMAN, M., & SCHEFFLER, I. Sex differences in emotional reaction to the

news. *J. abnorm. soc. Psychol.*, 1947, **42**, 476–479.

108. NORRIS, E. B., & GRANT, D. A. Eyelid conditioning as affected by verbally induced inhibitory set and counter reinforcement. *Amer. J. Psychol.*, 1948, **61**, 37–49.
109. NORTH, A. J. Performance during an extended series of discrimination reversals. *J. comp. physiol. Psychol.*, 1950, **43**, 461–470.
110. PORTER, P. B., STONE, C. P., & ERIKSEN, C. W. Learning ability in rats given electroconvulsive shocks in late infancy. *J. comp. physiol. Psychol.*, 1948, **41**, 423–431.
111. POSTMAN, L. Choice behavior and the process of recognition. *Amer. J. Psychol.*, 1950, **63**, 576–583.
112. POSTMAN, L., & JENKINS, W. O. An experimental analysis of set in rote learning: the interaction of learning instruction and retention performance. *J. exp. Psychol.*, 1948, **38**, 683–689.
113. POSTMAN, L., JENKINS, W. O., & POSTMAN, D. L. An experimental comparison of active recall and recognition. *Amer. J. Psychol.*, 1948, **61**, 511–519.
114. POSTMAN, L., & BRUNER, J. S. Multiplicity of set as a determinant of perceptual behavior. *J. exp. Psychol.*, 1949, **39**, 369–377.
115. PRENTICE, W. C. H. The relation of distance to the apparent size of figural after-effects. *Amer. J. Psychol.*, 1947, **60**, 617–623.
116. PRESTON, M. G., SPIERS, A., & TRASOFF, J. On certain conditions controlling realism and irrealism of aspirations. *J. exp. Psychol.*, 1947, **37**, 48–58.
117. REYNOLDS, B. Resistance to extinction as a function of the amount of reinforcement during acquisition. *J. exp. Psychol.*, 1950, **40**, 46–52.
118. RYAN, T. A., COTTRELL, C. L., & BITTERMAN, M. S. Muscular tension as an index of effort: the effect of glare and other disturbances in visual work. *Amer. J. Psychol.*, 1950, **63**, 317–341.
119. SATTERTHWAITE, F. E. Synthesis of variance. *Psychometrika*, 1941, **6**, 309–316.
120. SATTERTHWAITE, F. E. An approximate distribution of estimates of variance components. *Biometrics*, 1946, **2**, 110–114.
121. SCHOENFELD, W. N. The treatment of multiple entries in analysis of variance with three criteria of classification. *Amer. J. Psychol.*, 1944, **57**, 500–508.
122. SIEGEL, P. S., & STUCKEY, H. L. The diurnal course of water and food intake in the normal mature rat. *J. comp. physiol. Psychol.*, 1947, **40**, 365–370.
123. SMITH, H. F. Analysis of variance with unequal but proportionate numbers of observations in the sub-classes of a two-way classification. *Biometrics*, 1951, **7**, 70–74.
124. SNEDECOR, G. W. *Statistical methods.* Ames, Iowa: Collegiate Press, 1946.
125. SOLOMON, R. The role of effort in the performance of a distance discrimination. *J. exp. Psychol.*, 1949, **39**, 73–83.
126. TANG, P. C. The power function of the analysis of variance tests with tables and illustrations of their use. *Stat. res. Mem.*, 1938, **2**, 126–149.
127. THOMSON, G. H. The use of the Latin square in designing educational experiments. *Brit. J. educ. Psychol.*, 1941, **11**, 135–137.
128. TSAO, F. General solution of the analysis of variance and covariance in the case of unequal or disproportionate numbers of observations in the subclasses. *Psychometrika*, 1946, **11**, 107–128.
129. TUKEY, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, **5**, 99–114.
130. WEBB, W. B. The motivational aspect of an irrelevant drive in the behavior of the white rat. *J. exp. Psychol.*, 1949, **39**, 1–14.
131. WILSON, J. T. The formation and retention of remote associations in rote learning. *J. exp. Psychol.*, 1949, **39**, 830–838.
132. YATES, F. The analysis of replicated experiments when the field results are incomplete. *Emp. J. exp. Agric.*, 1933, **1**, 129–142.
133. YOUDEN, W. J. The Fisherian revolution in methods of experimentation. *J. Amer. statist. Ass.*, 1951, **46**, 47–50.

Received February 1, 1952.

THE RESPONSE TO COLOR AND EGO FUNCTIONS[1]

ROBERT H. FORTIER
Western Reserve University[2]

It is the intent of this paper, first, to present a theory—largely derived from Schachtel (**77**) and Rickers-Ovsiankina (**74**)—and second, to examine a number of studies involving color in an effort to substantiate and clarify the theory. The reader will soon observe that the paper is heavily weighted with material from Rorschach inkblot studies. This weighting is explained by two somewhat mutually dependent factors: color, as well as a number of other variables, has been systematically treated in the Rorschach test, and there have been a multitude of papers written in which the Rorschach test was featured. Whenever possible, parallel studies which use other techniques employing color will be introduced.

Derivation of a Theory

Schachtel

Schachtel feels that the experience of color and the experience of affect have two important characteristics in common: ". . . the passivity of the subject, and the immediacy of the relation object-subject" (**77**, p. 399). Two examples from Schachtel may clarify this analogy.

1. An individual enters a room in which there are two designs. On one wall is a large blob of color. On the opposite wall is a large design in black and white. The blob of color is immediately perceived, almost without conscious attention. The individual is aware only of color. The design in black and white requires directed attention before it can be perceived.

2. An individual becomes angered. He strikes out blindly at his antagonist, without regard for the consequences of his act. He is aware only of his anger and an object upon which to vent this anger.

The two examples are extremes. One may imagine more moderate behaviors in the two situations described above. An individual entering the room immediately perceives the blob of color, but his perception also encompasses the contour, and some analysis of its shape occurs. An individual becomes extremely angry but acts upon the situation in such a manner that the tension produced by the anger is reduced without violence being done either to the stimulus of the anger or to the individual himself.

It is a primary task of the ego to control and direct affective reactions, whether produced by drives originating from without the ego or from within it (**40**). The individuals in the two situations described by Schachtel were completely passive. They were literally swept away by their experiences. Their egos, which should have channeled and controlled the aroused experiences, failed in their tasks. The behaviors of the individuals in the situations described by the writer may have retained some elements of

[1] This paper represents a slight modification of the first chapter of a Ph.D. dissertation, A Study of the Relation of the Response to Color and Some Personality Functions, submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy, Western Reserve University, 1952. The writer gives his sincere thanks to Dr. Calvin S. Hall whose astute and acute criticisms contributed much to the formulation of this paper.

[2] At the time this paper was submitted, the writer was a USPHS postdoctoral research fellow at the Roscoe B. Jackson Memorial Laboratory. At time of publication, the writer is a postdoctoral training fellow in clinical psychology at the VA Hospital, Northampton, Massachusetts.

passivity, but this passivity was considerably moderated by ego control. It is evident that passivity does not refer to the overt behavior of the individual, but only to the relation between his affective drives and his ego. Schachtel points out that the affective experience is a conscious one, regardless of the passivity of the ego. Where there are no affects, there is no consciousness of drives.

Rickers-Ovsiankina

Rickers-Ovsiankina, after reviewing the literature (**74**), concludes that an individual's response to color can give considerable insight into the degree of permeability of his ego. That individual whose ego is responsive to the outside world will respond to color. As the degree of permeability increases—that is, as the boundary between the ego and the outside world lessens in strength—the individual will respond more to color per se.

Both Schachtel and Rickers-Ovsiankina stress the fact that the extratensive individual is responding to outside stimuli and that there is a lessening in the spontaneity of the individual. However, one misses an important point if he concentrates upon this particular elaboration in Rickers-Ovsiankina's paper. Schachtel feels that the individual who responds to color per se not only possesses an ego which readily responds to the outside environment but which also is less capable of exerting control upon affective drives having internal origin. He feels then that the permeability of the ego is a two-way affair, for there is also a more direct release of affective drive upon the external environment.

The Concept of the Egocentric Individual

Let us diverge for a moment to consider another aspect of the individual who responds to color per se. Such a divergence will serve the dual purpose of clarifying the points presented thus far, and of facilitating interpretation of certain of the studies which will be presented later.

The individual who responds to color per se has been called egocentric by a number of authors (**11**, **16**, **41**, **42**, **54**, **68**). In the light of the preceding discussion, just what does "egocentricity" mean? Warren (**88**) defines the term "egocentric" as follows: "disposed to dwell on oneself and to view every situation from a personal angle" (**88**, p. 89). As a synonym, he gives the term "self-centered." Do Schachtel and Rickers-Ovsiankina actually consider the individual who responds to relatively undifferentiated color "egocentric"? Under Schachtel's scheme the individual adopting this mode of color response may behave in one of two ways, or both.

The individual will adapt to his environment, behaving as it dictates. If an environmental configuration directs action in one way, he will act in that way. If action is directed another way, then the individual will again modify his action to conform to environmental pressure. This individual clearly cannot be called "self-centered," for he is most likely to perform in the manner in which others wish him to perform. But Schachtel further feels that the individual responding to relatively undifferentiated color may also release his affective drives in a relatively undifferentiated manner upon the environment without regard for that environment. It is this latter behavior which might best be called "egocentric." Nevertheless, Schachtel makes it clear that an individual who behaves in this manner is not channeling his affective drives. There

is a more or less direct interchange between his affective drives and the environment. The individual's behavior, therefore, rather than being egocentric, is relatively removed from the control of the ego. The ego plays a relatively unimportant role as to the object upon which the affective charge is released and the manner of its release. It is certainly far from a deliberate (that is, ego in origin) attempt to ignore the feelings of others that results in the sometimes antisocial and inconsiderate behavior of this individual. It is the result of a fundamental incapacity of the ego to direct and control the affective charge in a realistic manner.

Statement of a Theory

The experience of affect and the experience of color are quite comparable. Thus one may examine the less obvious of the two, affect, by the response to the more obvious, color. The experience of affect is passive. The degree of passivity is determined by the degree of control exerted upon the affective charge by the ego. That individual who responds to relatively undifferentiated color possesses an ego which is less able to control and channel affective charge. Such an individual lacks spontaneity of action and readily adopts the color of his environment. It is a logical corollary that an individual responding to relatively undifferentiated color may release affective charge upon the environment in a more or less undifferentiated manner. Further, his perception of affective charge in others may also be relatively undifferentiated. That is, the individual may sense very acutely the presence of affect in others without being able to differentiate it, to identify its nature. Thus, when someone becomes angry with him, he may only be aware of the existence of a powerful affective state in his antagonist. He may not know what is the nature of this state. However, the particular course of action followed by the individual depends upon his total personality configuration.[3]

The Studies

A theory, no matter what its logical integrity, must be tested by data set forth in a variety of studies. Rickers-Ovsiankina in her paper (**74**) to which reference is made above, reviews a number of articles and the present writer does not intend to duplicate her bibliography to any great extent. The studies included here vary from those dealing with normal individuals and their development through those concerned with individuals suffering from organic brain damage.

Since a number of studies to which reference will be made later concern the Rorschach test, brief mention will be made of the treatment of color by users of the test. Those desiring a comprehensive exposition of the treatment of color by Rorschach investigators may refer to any of the standard texts (**11**, **12**, **16**, **54**, etc.) or to the normative studies of Hertz (**41**, **42**). Suffice it to say that the scoring of a response to color depends upon the degree of structure imparted to

[3] The reader who refers to the original manuscripts by Schachtel and Rickers-Ovsiankina will see that for the most part the statement presented here is simply a more succinct and perhaps clearer presentation of some of the ideas formulated in the two papers. The suggestion that a blunting of the perception of affective drives in others accompanies the response to relatively undifferentiated color may be considered the most important addition. The writer feels that at this stage of knowledge, little may be gained from further additions to the theory, but that the theory and the manner in which it is used in the present paper can serve to promote more rigorous investigations, in that way contributing to an advancement of the theory.

the color or the degree of integration of color with form. An undifferentiated or unstructured color response is one determined solely by the color of the blot. This is scored as C. A response which is determined principally by the color but which is structured to some extent is scored as CF. A response in which the color is quite integrated with form is scored as FC. The nature of the affective experience which is represented by each of these scoring categories is apparent from what has been said above. The three color factors have been assigned numerical weights as follows: FC, .05; CF, 1.0; C, 1.5. These weights were originally suggested by Rorschach, and have been used extensively. Rorschach's belief was that, since C represented a more powerful and uncontrolled affective drive, it should have the greatest weighting and FC the least. Weighting finds applicability particularly in determining the so called stability or control ratio, calculated by the formula

$$FC-(CF+C).$$

The Normal Picture

The criteria for normalcy may be statistical, may be based upon psychological theory and knowledge of dynamics, or may be philosophical. It is not the purpose of this paper to delve deeply into the dynamics or total configuration of any group. Therefore, for the criteria of normalcy (in relation to the affective aspect), the following should suffice.

First, it seems logical that in our society an individual must have affective drives and affective relationships with others, and must be relatively in control of these drives and relationships. That is, the direction and manner of release of affective charge must be ego-controlled. He must also be able to interpret and integrate the affective behavior of others. Second, the world must not be so firmly fixed and structured in his ego that he can not be moved or partially influenced by the particular configuration of the environment. To a certain extent, his affective behavior towards others should not continually require meditation and deliberation. Third, affective reactions directed towards the environment by the individual cannot be overly gross and undifferentiated, nor can his perception of and reaction to the affective behavior of others be undifferentiated or gross. The type of color response on the Rorschach Test representing each of the above delineations is obvious: the first by FC; the second by CF; the third by C.

The Normal Adult

Rorschach findings. What is found by Rorschach examination of normal individuals? There are several sources of information, among them the standard texts and normative studies by Hertz already cited. Klopfer and Kelley (**54**) suggest that the normal adult should give some color responses, but that the sum weight of C and CF responses should not be higher than the sum weight of FC responses. They feel that a crude C response, one which is not descriptive or symbolic, is a pathological sign. (This view is not shared by Hertz.) Beck (**11, 12**) gives no specific norms for any group, but presents the psychological significance of the various types of color responses and suggests some individuals who would give them. The sign of the healthy individual is FC. Such an individual is mature and can establish affective relations with other individuals. This individual may give a number of, or a few CF responses, but the sum weight of the FC responses should approximately equal or exceed

the $CF+C$ sum weight. C is suggestive of regression.

In a major normative study by Hertz and Baker (**42**) it was found that an average of 3.7 color responses is given by 15-year-old boys and girls. No averages are given for *CF* or *C* responses. However, the range for *CF* responses is 0–2; the range for *C* responses is 0–1. Thus, one *C* response would not be considered pathological. The average sum weight of color responses is 2.8. Balance in favor of *FC* is indicated by the weighted ratio $FC-(CF+C)$, which is $+0.54$. The implication is clear that Hertz and Baker would expect normal adults to have at least as balanced a ratio in favor of *FC*, or a higher one. Steinzor (**84**), in testing a presumably normal college group, found a balance in favor of *FC*. In a healthy individual, a sign of integration is a balance in favor of *FC*, with a total color weight of 3.0.

Beck, *et al.* report a normative study based on a fairly large number of adults. Concerning the color factors, they conclude:

> Of most interest is the weighting in the direction of *CF;* and *FC*, in that order; and the comparatively small instance of undiluted *C*. The population of which this sample is representative may, therefore, in respect to affectivity, be described as having made some progress towards maturity and towards capacity for social rapport. Yet they are slightly more labile than fully stabilized. On the other hand, the quantity of infantile egocentricity is relatively small . . . unstable, easily excited, but resisting undisciplined violence (**13**, p. 259).

The reader may wonder whether the theory has after all been formulated incorrectly, with this sudden finding of a predominance of *CF* over *FC*. However, there is another explanation which, to the writer, seems rather feasible. It may be, as Beck suggests later in the study, that the dynamic configuration of the individual and the society has changed. The color factors reflect this change. A clarification may result if one attributes to these findings the psychological significance formulated by the theory stated in the present paper, without allusions to maturity or infantile reactions. The presence of *FC* indicates a capacity for ego-controlled affectivity and the capacity to integrate and interpret the affective behavior of others. The excess of *CF* over *FC* indicates that the individual may have an immediate reaction to the environment and may be considerably influenced by it. The very small instance of *C* is a countersign against gross and undifferentiated affective behavior and perception. In other words, although the present adult may react and respond far more readily to his environment than he once did—if the earlier diagnoses were correct—he still has the capacity for constructive affective behavior.

Studies involving painting. The majority of studies involving easel and fingerpainting had children as subjects. However, Waehner (**87**), from his work with college students, developed several indices in regard to emotional balance, control, compulsion, constriction, etc. In constructing these indices, Waehner deliberately paralleled to a considerable extent certain Rorschach procedures. Superior emotional balance is reflected by a color variety of three to six, and a relationship of color to form of $5C:4F$. Constriction is indicated by a small variety of or no color.

Napoli (**65, 66**) will not be discussed here since he is more concerned with the particular hues rather than color in general. Those interested in a broad survey of paint-

ing are referred to a recent article by Precker (**72**).

The Mosaic Test. This instrument has been gaining in interest among individuals with varied approaches to study of personality, but as yet few articles have been published. Wertham and Golden (**92**), while more concerned with the forms of the designs reproduced, expect normal individuals to produce designs harmonious in color and distinct in configuration. Diamond and Schmale (**26**) say that normals may have a very wide range in the use of color from primitive and crude designs in color to extremely artistic use of color. In comparing data obtained from the Mosaic Test with that obtained from the Rorschach test, the authors conclude that there is a tremendous discrepancy between the results obtained by the two instruments. This discrepancy and some possible theory underlying the varying performances obtained with the two instruments will be elaborated when psychotic modes of adjustment are discussed later in this paper. Lowenfield (**60**) confirms Diamond and Schmale's (**26**) findings on normals.

A Genetic Approach

Many investigators feel that a correlate of increasing chronological age is increasing emotional control. From an examination of genetic studies, therefore, one should be able to determine, first, how the individual performs at different ages, and second, what is the psychological significance of this performance. Jersild reviews a number of studies of emotional development and concludes: "The data now available from direct observation or experimental study do not provide the basis for a systematic account of normal and immature emotional behavior at various age levels" (**45**, p. 760). Therefore, what material can be cited here is admittedly sketchy and incomplete.

Pratt, Nelson, and Sun (in **45**), as a result of their studies of neonates and slightly older children, stress the point that "generalized reactions predominate over specific reactions in early childhood and the fact that distinctive patterns are difficult to detect." Taylor (in **45**) also stresses the undifferentiated nature of emotional reactions in a study of children aged ***one to twelve days***. Sherman (in **45**) arrives at a similar conclusion and expresses the belief that "with the passage of time the child's behavior becomes increasingly differentiated and adaptive." Jersild (**45**) points out that the manifest emotional behavior may change through the changing nature of the emotional problems with which the child is confronted.

Gesell (**32**) comments briefly on the problem of emotional development. We may compare what he says with what may be inferred from other sources.

> The three year old attempts to conform and please, "as though he were sensitive to the demands of the culture." Suggestions are accepted more readily. He may prefer the companionship of other children but as yet is incapable of verbalizing his desires. He can play with children for a while, but may suddenly attack them. He is at least somewhat susceptible to social suggestion. He studies the facial expressions of individuals in his environment and attempts to interpret them. "He is capable of sympathy." Similar sketches are drawn for the four- and five-year-old. The progress made by the three-year-old child is continued, but the pattern may change slightly. "Three has a conforming mind. Four has a lively mind. Three is assentive; four assertive." Five shows much more definiteness, concreteness. Gesell calls this age a plateau.

The Rorschach findings. Unfortunately, the sources of information are

again meager. Klopfer and Marguiles (**55**) made a study of children aged two through six years. They present percentages of children in each age group who use the various color scoring categories, and the average number of such responses given by each child. The results on the color factors are reproduced in Table 1. The authors' findings that at the age of six *FC* dominated the types of color responses should be particularly noted.

A second group of children, aged three through seven years, was studied several years later by Ford (**30**). To facilitate interpretation and comparison with the results obtained by Klopfer and Marguiles, a summary of her results on the color factors is presented in Table 1.

TABLE 1

PERCENTAGE OF CHILDREN OF DIFFERENT AGE LEVELS GIVING COLOR RESPONSES AND THE AVERAGE NUMBER OF COLOR RESPONSES GIVEN

Ages	*FC*	*CF*	*C**	*FC*	*CF*	*C*	*Cn†*
			(*After Klopfer*)				
2	12	12	29	.29	.23	.59	.23
3	37	24	46	.87	.41	1.04	.48
4	27	45	30	.88	1.24	.54	.09
5	54	67	41	1.29	1.44	.75	—
6	65	43	30	2.08	.78	.69	—
			(*After Ford*)				
3	28	32	12	.5	.5	.2	.7
4	44	56	15	.8	.8	.2	.7
5	56	36	36	.8	.5	.8	.4
6	57	61	35	.7	1.0	.5	0
7	52	57	43	1.1	1.2	.7	0

* In terms of per cent.
† In terms of average number.

A principal source of discrepancy, at least in reporting the percentages of children giving the responses, is derived from the fact that Ford very carefully distinguished between pure *C* responses and color naming (*Cn*), that is, where an individual simply gave the name of the color. Klopfer and Marguiles made no such distinction in reporting percentages.

An explanation of the diverging results obtained by Ford and Klopfer and Marguiles is difficult to find. Both studies were based upon children who could be expected to have comparable socioeconomic backgrounds. One source of variance may lie in the particular scoring bias of the different investigators. The records used by Klopfer and Marguiles were submitted by a number of different investigators, however, and several other investigators confirmed the findings by Ford. It may be that a difference in time—the year in which the investigations were made—is a contributing factor. The study by Klopfer and Marguiles was reported in 1941; the study by Ford in 1946. Obviously both investigations took much time to prepare. If it can be determined that the study by Klopfer and Marguiles antedated by any great period of time that by Ford, this difference, when interpreted in the light of the recent study by Beck (**13**), may point to a fairly rapid altering in the dynamic configuration of the individual and the society. This is particularly true of the discrepancy obtained in regard to the dominance of *FC*.[4]

Studies involving painting. The text by Alschuler and Hattwick (**3**) undoubtedly represents the most com-

[4] The mounting percentages of children giving color responses associated with increasing age are readily explained if one recalls Schachtel's (**77**) injunction that affects are the conscious representations of drives. Obviously, a child becomes more aware of his drives with increasing age. Therefore, a very young child giving many color responses is showing signs of "precocity" rather than "infantilism."

plete and recent major work concerned with painting and children. The authors' findings tend to confirm Rorschach test results with children. Children of three or four are quite interested in color and tend to use it without great regard for form. With increasing age, the color tends to become more and more integrated with form. In studying groups of children, one of which uses a great deal of color and one of which is more concerned with form, the authors found that children concerned with form were more self-controlled, more concerned with external stimuli, and had a higher frequency of reasoned (in contrast to impulsive) behavior than those using much color.

Epstein and Schwartz (**29**) found that the number of colors used reflects the emotional development of the child; those using under four colors having poor emotional development, lack of drive, and perhaps constriction. Overcontrol or retarded development is indicated by a predominance of form over color. That interest in and use of color declines after the child reaches a certain age was confirmed by Blum and Dragowitz (**15**).

Thus it is seen that although an integration of form with color in painting—at least to a certain extent—may be expected from children who have acquired some facility with emotional control, there is a certain time lag involved as to when such integration occurs. A suggestion is that the two methods of studying emotional development—the Rorschach test and painting—may measure different aspects of this development. Since the writer feels that, to a considerable extent, the Mosaic Test and easel painting are functionally comparable, a discussion of the discrepancies obtained through use of the Mosaic Test and the Rorschach test will be discussed later as already indicated.

Contribution of the Genetic Approach and the Adult Picture

The task is now to re-examine the psychological significance of the color factors in the light of the evidence presented in these two sections. Beck (**12**), in elaborating the psychological significance of the pure *C* response, says: "This is the reaction mode of the infant, who does what he pleases—screams, demands food, kicks, voids without regard to time and place. Response to feelings is exclusive and instant" (p. 30). Jersild (**45**) holds such reactions as described by Beck as typical of very young children, neonates, and children up until perhaps the age of two. The studies reviewed by Jersild (**45**) definitely point out that emotional control rapidly increases with increasing age. A neonate obviously cannot be given a Rorschach test. If Ford's results (and those of investigators reporting similar findings) may be accepted as at least typical of a certain class of children, it is seen that three-year-old children as a group gave but 0.2 *C* responses. It then becomes difficult to see upon what basis an interpretation such as Beck's is founded. Gesell (**32**) paints a picture of the three-year-old as an individual who has gained considerably in emotional control, but who is still markedly dependent upon the desires and wishes of those in his environment. He is also prone *upon occasions* to attack quickly and violently those individuals around him. He is interested in studying and trying to interpret the facial expressions of those around him.

From this picture, something comparable to Ford's results could have

been expected. Of the three-year-old children, 28 per cent gave *FC* responses. *CF* responses predominate, while *C* responses are last in both per cent and average number. The finding that *CF* predominates at this age level, coupled with the statements that the three-year-old child is "assentive" and is ". . . sensitive to the demands of the culture . . ." (**32**, p. 36) corroborates the interpretation of the *CF* response presented by the present writer. The statement by Gesell (**32**) that the five-year level constitutes a plateau coincides quite well with the finding by Ford that it is only at this age level that a predominance of *FC* over *CF* is found. The significance of the *C* response also seems to mean what it was sketched as meaning in the introductory theory—affective charge not controlled or directed by the ego.

The significance of the *Cn* (color naming) response is still to be determined. Anyone observing the relations of young children and parents will find, during a certain phase in the child's development, parents pointing to different colored objects, regardless of their shapes, saying, "this is green, this is yellow," etc. It therefore seems logical that, when a young child is presented with a new type of game, he will point out certain areas on the card and say, "this is green, this is yellow," etc. Or, if one assumes that the hypothesized relation between affect and color holds also in this situation, another interpretation is possible. Gesell (**32**) points out that the child studies the facial expression of those around him and tries to interpret them. The interpretation may be of the nature of nosology: that is, the attempt may be to identify, without the ability to act upon the interpretation. Then, may not the child, after looking at his father, say to himself, "he is angry," or "he is sad," without having the ability to act constructively upon his identification?

No contradiction is found between the psychological significance that is attributed to the color factors when used by the child and the psychological significance that is attributed to the color factors when used by the adult. The writer feels that if one is successful in defining the psychological significance of a variable, and this significance remains constant whether the variable is used by an adult or a child, it is better to use this definition than to attempt to define the variable in terms of one or the other chronological referents.

Nonpsychopathological Deviations

The Institutionalized Child

Turning from normal individuals raised and living in a normal environment, attention can be focused upon those individuals—more specifically, children—who have spent the greater portion of their lives in institutions. As a result of a study by Goldfarb (**34**), it was found that more such children give the pure *C* response, and that more of them exhibit the unbalanced ratio of sum weight *CF*+*C* greater than *FC* weight. Goldfarb suggests that this is an indication of the lessening of rational control and a greater emotional immaturity.

Goldfarb and Klopfer continue this analysis, concluding that:

> The institutional group thus shows deficiencies in rational control, in more abstract forms of thinking, in drive for intellectual and social attainments, and in emotional maturity. In a group with such psychological tendencies one would, of course, expect problems involving restlessness, inability to concentrate, and poor adjustment. In addition, all of the

above listed Rorschach trends among the "institution" children are associated with an *air of passivity* (italics added). In other words, the children of this group give little of themselves though superficially they are adjusting to reality requirements (**36**, p. 93).

The writer feels it significant that Goldfarb and Klopfer find an air of passivity in institutionalized children, for were one to attribute the usual significance to an unbalanced $CF+C$ ratio, a far from passive attitude would be expected—vigorously negativistic, impulsive, willful, etc. What this ratio actually seems to suggest here is a greater inclination to be moved and swayed by the environmental configuration in which the children find themselves. It suggests a more undifferentiated emotional approach to the environment, possibly a rather diffuse emotional reaction towards everyone with whom the children come into contact. Such behavior would logically follow from the nature of the institutions in which the children find themselves forced to live. Their behavior is in virtually every respect governed by more or less impersonal rules and regulations. Their relations with adults are limited simply because there are so few adults in the institutions that personal contact is quite difficult. The suggestion follows naturally that for one to have sufficient intellectual control of affectivity one must have the opportunity to learn and develop this capacity. It is probable, then, that the nature of one's affective life is dependent, above native endowment, perhaps, upon the ability and opportunity to learn.

The Delinquent

In regard to this last postulate—affective control and opportunity for learning—the findings of those investigators concerned with juvenile delinquents may have some bearing. "Burt holds that marked emotionality is the most frequent and most influential of all the psychological characteristics of the delinquent" (**27**, p. 129). This statement sums up succinctly an attitude towards the genesis of delinquency which prevailed for a considerable period of time, and perhaps prevails in certain quarters now, judging from the studies still concerned with the relation between emotionality and delinquency.

Rorschach findings. The Rorschach findings may prove rather startling to those investigators still adhering to the classical theory of delinquency quoted above. Endacott (**28**) in a study of 100 delinquent boys—average age, 14 years—found a restriction of color, lower FC, and lower CF when his results were compared to those of other investigators. The normative study by Hertz and Baker (**42**) suggests that a *sum C* weight of one to one and one-half points higher could be expected from boys of this age. Boynton and Walsworth (**18**) report a study of 47 delinquent vocational school (reform school) girls of approximately high school age. The authors compared the results obtained from the Rorschach protocols of the delinquent girls with those obtained from girls attending a high school located in a favorable section of the same town. In regard to color, the delinquent girls scored lower than the high school girls in all respects. The so-called impulsivity ratio was more in favor of $CF+C$ in the high school group than in the delinquent group. A number of earlier studies reporting excessive emotionality in delinquents were reviewed by Schmidl (**79**) and criticized because of inadequate sampling and other factors. He points out the suggestion by

Beck that delinquents can be either extratensive or introversive.

These later Rorschach findings suggest that some cause other than marked emotionality must be postulated to explain the antisocial behavior of delinquents. There is a suggestion that the inability to establish rapport (or lack of ego-controlled affective charges) cannot be accepted as a general factor in explaining delinquency. Boynton and Walsworth (**18**) feel that one should be quite careful in using personality aberrations as explanations for delinquent behavior. Endacott sums up his findings by saying, "these are marks of a rigid, stiff-geared sort of personality that has been created to withstand strong pressures and frustrations" (**28**).

The implication for affect or color theory is clear. If it can be shown that a group having a somewhat lower use of color and a more stable color ratio than "normal" individuals indulges in strong, "self-centered," antisocial behavior, it is a logical deduction that it is not the affective relationship with the environment which should be postulated as a causative factor. There is a further indication that the presence of ego-controlled affectivity or "capacity for rapport" suggests nothing more concerning the individual than that he can direct and control his affective charges and can interpret and integrate the affective behavior of others. "Capacity for rapport" suggests nothing concerning the content of the behavior of the individual. The delinquent, it would appear, is acting in accordance with his picture of the reality, i.e., in accordance with his ego. A logical corollary of the above deduction is this: where antisocial, self-centered behavior appears concomitantly with extratensive, unbalanced use of color, one should look beyond affect for an explanation of this behavior.

A study using fingerpainting. It has often happened that where one method failed to provide insight into a particular problem, another method has succeeded at least partially. The method of approach to and the products of fingerpaintings of delinquent and high school youths were compared by Phillips and Stromberg (**69**). While a number of significant differences are reported in their study, only one need be discussed here. Thirty-six per cent of the high school group used only one color on the first performance. Sixty-four per cent of the delinquents used only one color on the first performance. This difference is not quite significant at the .05 level. However, on the second performance, 4 per cent of the high school students used only one color, while 60 per cent of the delinquents continued to use only one color. The result of this comparison is highly significant statistically.

If it can be assumed that one's handling of color is indicative of his affective life, the nondelinquent, and the delinquent to a greater degree, might here be showing a certain amount of shock when confronted with a new—and perhaps affective—situation, and thus respond in a somewhat stereotyped manner. However, the nondelinquent shows a considerable degree of recoverability and a capacity for a wide variety of response. The delinquent, on the other hand, continues to show a stereotyped reaction. When one recalls a deduction made from the evidence reviewed concerning institutionalized children, it is possible that the environment in which delinquents live does not make it possible for them to learn a variety of emotional reactions. Thus, while it may not be any fundamental lack of capacity for emotional

rapport or ego control of affect which contributes to their antisocial behavior, it may be an inability of the delinquents to vary their emotional response. This suggestion accords with the reduced use of color in general on the Rorschach test and the reduced use of *CF* by delinquents.

Psychopathological Deviations

Investigators have found that considerable insight may be obtained into certain dynamic relationships and functions by studying individuals exhibiting more or less psychopathological reactions. It is reasonable to expect that a comparable result may be obtained here by reviewing the behavior of such individuals towards color.

Alcoholics

Using the Rorschach test, Billig and Sullivan (**14**) found that the affective picture presented in the use of color by alcoholics is of considerable prognostic value. In reference to those alcoholics who over a period of time showed the least favorable prognosis, the authors conclude: "In 80% of the cases factors indicating impulsive emotional behavior are stronger than those expressing smooth adjustment to environmental influences" (**14**, p. 124). However, if their table is accurate, the impulsive use of color is reflected primarily by the use of *CF* rather than *C*. There is a marked reduction in the appearance of *FC*, and consequently the color ratio is unbalanced in favor of *CF*. The authors feel that their results confirm a previous study by Bowman and Jellinek, who made the statement that "the chronic alcoholic shows a comparatively weak restraint, poor mental poise and stability, difficulties in controlling his mood swings and desires, combined with a lack of attention" (**14**, p. 124).

It is immediately apparent that the statement cited from the paper by Bowman and Jellinek presents an interpretation which is far more comparable to that which would be made by the present writer (drawing upon the theoretical outline sketched) on the basis of the evidence presented by Billig and Sullivan than the one actually made by the latter authors. The comparative lack of *FC* responses among the alchoholics with a poor prognosis, coupled with the overabundance of *CF* responses, suggest that these alcoholics are particularly affected by the environmental configuration in which they are immersed and are lacking the capacity to integrate and control the affective drives which are therefore readily aroused in them. The appearance of pure *C* in these cases would simply add a more unfavorable touch by suggesting relatively undifferentiated and diffuse emotional reactions and interpretations.

Enuretics

Enuretic children under ten years of age were shown by Goldfarb to have a high *sum C* total "with a conspicuous excess of *CF* and uncontrolled *C* responses. Emotional development at a primitive, infantile, impulsive level is suggested" (**33**, p. 30). Specific figures are not given. However, a glance at the studies by Ford (**30**) and Swift (**86**) show that, when color naming is excluded, pure *C* is not a particularly frequent type of response. To be sure, *CF* outweighs *FC*, but some *FC* does appear. A high frequency of *C* and *CF* responses does not seem to be typical of children of the ages studied so far, and considerable difficulty would be encountered in the task of determining whether such a reaction is typical of infants in the technical sense of the term, i.e., from birth to two years.

For one to call this excessive use of unbalanced color *infantile* is to use an analogy which is dubious and which may never be confirmed. Even were the analogy correct, to say that something is infantile is not in itself particularly expressive because little concerning the dynamics of the function involved is suggested. What this mode of usage of color by enuretics suggests (as has been shown in other cases) is a considerable degree of influence by the environment, with a relatively diffuse reaction towards it, and a reduction in the ego control of affect. Study of the ego of the enuretic might be more revealing of the dynamics of this particular dysfunction, i.e., enuresis, than analysis of the affective factors alone. Then, one may infer that this undifferentiated type of reaction and excessive influence of the environment contribute to the development of the symptom. Goldfarb's study does contribute to our knowledge of the ego content of the enuretic. Only the interpretation of the color factors is in question.

The Hysteric

Schafer (**78**) suggests that a characteristic of the hysteric is a predominance of $CF+C$ over FC in the Rorschach test. A further characterization is a "minimization of active and independent ideation as a means of coping with problems" (**78**, p. 33). Such characterizations tend to substantiate the theory that a less integrated color response suggests a greater susceptibility to environmental influence and a lessening control of one's affect.

The problem of egocentricity enters the picture. The present writer feels, as he earlier proposed, that the presence of relatively uncontrolled color is neither an indication of nor a countersign against egocentricity as it appears in hysterics and others. He feels that it would be much more feasible to regard the behavior of the hysteric which leads others to call him egocentric as a quality of the ego content, or picture of reality, of the hysteric rather than of his affective life.

Schizophrenic Adjustment and the Rorschach Test

Among the psychotic modes of adjustment, that of schizophrenia has attracted the most attention. No detailed review of the dynamic picture of the schizophrenic will be attempted since the task would be complicated by the belief held by many investigators that schizophrenia is not a single disease entity. Beck (**9**, **11**, **12**), Kelley and Klopfer (**49**), Klopfer and Kelley (**54**), Rickers-Ovsiankina (**73**), Kisker (**52**), Stern and Malloy (**85**), Kendig (**51**), and others, feel that a characteristic of schizophrenia is an overwhelming imbalance in color responses in the direction of $CF+C$. The total color weight may be large or small. As Beck (**9**) suggests, affect is not absent or even negligible in the schizophrenic as was once felt to be the case.

The color factors indicate first, according to the present writer's introductory scheme, that the schizophrenic is considerably influenced by the environmental configuration in which he finds himself; second, that his perception of affective life in others and his affective reaction to the environment are undifferentiated and gross. The inappropriate affective reaction frequently noted in the schizophrenic is explained by the finding of pure C in the record. If the schizophrenic cannot interpret correctly the nature of the affective situation with which he is confronted, this, coupled with the fact that he has

little ego control of his affectivity, certainly would make it rather a coincidence if appropriate emotional reactions did result.

To postulate that the schizophrenic is considerably influenced by the environmental configuration may deviate somewhat from the typical concept of the schizophrenic as separated from reality. The mere fact that the schizophrenic is considerably influenced by the environment does not indicate that his reactions to the environment will be realistic. The content of his behavior is determined by the content of his ego. It is agreed that the ego of the schizophrenic contains far from a realistic conception of the environment. Therefore, this environmental influence operating upon a distorted picture of reality would only add to the confusion and bizarre reactions of the schizophrenic. The susceptibility to environmental influence might explain the finding by several clinicians that a great deal of what occurs around the schizophrenic in a catatonic stupor is frequently remembered by the schizophrenic when he recovers from the stupor.

Schizophrenic Adjustment and the Mosaic Test: The Suggestion of a Theory

Diamond and Schmale's (26) investigation of schizophrenia by use of the Mosaic Test, concerned with the dynamics of the schizophrenic, theory of the Mosaic Test, and affect-color theory, may prove to be of considerable value. The authors found that the schizophrenic completely disregarded the color of the pieces in the construction of his design.

> The color defects of the schizophrenic deserve special discussion. Color rejection or color disregard appears very early in this disease even though the personality and the Mosaic pattern are seemingly well-integrated. . . . The Mosaic pattern is exactly as if it were constructed by a totally color blind individual. It might be called a psychological color blindness. . . . It was very difficult to compare the color responses to the Rorschach and the Mosaic Tests in individual cases, and little consistency between the two were shown (p. 246).

An obvious source of this inconsistency with the two tests is in their basic dynamics. The Rorschach inkblots are often called unstructured, ambiguous, undefined. This is true, but only in a certain sense. They do exist; they do have very definite and unchanging configurations. No matter how the subject looks at the blots, there is no change in the actual shape of the blots. The Mosaic Test is completely different. Before the subject are a large number of little blocks of many different shapes and colors. He moves them around and can put them back together in a multitude of different ways. (It can be seen that the functioning of an individual in regard to easel- or fingerpainting is comparable to what it is with the Mosaic Test. The writer feels that the task presented by the Mosaic Test is more difficult than that presented by finger- or other painting.)

Many investigators feel that the task demanded of the individual confronted with the Rorschach test is essentially a creative one. The same can be said of the Mosaic Test. But is not the nature of the creative activity, the basic mechanism of the creative activity, extremely different from one test to the other? The creativity involved in the Rorschach test is exclusively an associational one. Upon being presented with an unchanging configuration, the individual is asked to call upon the content of his ego for a concept which corresponds more or less to the actual shape of the configuration. The individual is not asked to alter reality.

He changes nothing in the external environment nor does he create anything in it. The task involved in the Mosaic Test is also associative, but it is far more than that. Before one constructs a definite thing in the external environment, he has a more or less defined image of that thing in his mind. The nature and variety of what is brought forth depends, as it does in the Rorschach test, upon the content of the ego and its associative facility. But the development of the concept or image must be followed by an alteration in the external environment. The resulting product depends upon manipulative skill to some extent, but to a greater extent upon the individual's capacity to translate his more or less well-defined image into concrete reality.

How does this difference in the basic mechanisms involved in the two tests affect the product? The response to color only is of pertinence here. Consider the schizophrenic individual whose capacity for ego-controlled and ego-oriented affective charges is considerably reduced, and whose emotional responses are becoming increasingly gross and undifferentiated. On a piece of cardboard before him, he sees a blob of color. Not much capacity or skill is required for this individual to call forth a vague, structureless association. More often than not, if he succeeds in developing a structured response, the response will not fit the configuration confronting him.

In the Mosaic Test, for the individual to handle color properly—that is, if he is not to ignore it—he must first be able to visualize or conceive an image or configuration involving color and then reproduce this image by an alteration of the external environment. An individual whose capacity for ego-controlled affectivity is greatly reduced would have much difficulty in conceiving an affective situation and more in manipulating it in the external environment. The product of this individual would, of course, show not only a disregard of the color of the Mosaic blocks, but very poor form as well.

But an individual who is not quite sure of himself, who is becoming aware that something about his handling of affectively charged situations is not quite right, would, even if he could conceive a fairly adequate affectively charged image, hesitate to bring this image forth, to reproduce it in the external reality where it would be visible not only to himself but to others. Since Diamond and Schmale (**26**) found that even in the very early stages of schizophrenia color was ignored while the capacity to integrate form was relatively intact, the suggestion is strong that one of the very first signs of the schizophrenic process is an uncertainty, perhaps even consciously realized, that one is losing his capacity to handle affective situations.

The writer thinks it will be agreed that capacity to handle emotionally toned situations will vary among individuals who do possess facility to conceive and interpret such situations. If this is so, and if the writer's analysis of the process underlying the Mosaic Test is correct, it is not difficult to see why results obtained in one test are not comparable to those obtained in the other. Thus there are two aspects to affective situations: the ability to conceive and interpret them; and the ability to handle them in the external environment. The reasoning followed here tends to be confirmed by a study of institutionalized children conducted by Colm (**24**). These children also used color indiscriminately. It is obvious that institutionalized children have limited scope and opportunity

for learning to handle affective situations.

The Feebleminded Individual

Davidson and Klopfer (**25**), Kelley (**48**), Abel (**1**), and Werner (**90**) agree that on the Rorschach test the mental defective, although he may use less color, uses it in an unbalanced fashion, i.e., $CF+C$ greater than *FC*. Abel (**1**), in studying defectives showing the least inclination to succeed in school, found them to give more such unbalanced records than successful defectives. He concluded that the former are more susceptible to "stimulation from the external environment without adequate control of the situation."

Studies of Individuals Having Cerebral Disorders

Epileptics. Guirdham (**38**), Arluck (**6**), and Kelley (**50**) agree that the epileptic gives fewer color responses on the Rorschach test than the normal individual. However, these responses are definitely unbalanced in favor of $CF+C$. The suggestion is that the epileptic has little emotional communication with the environment, and that such communication as does occur is not under ego control, tending to be very gross and undifferentiated. It is of considerable significance that Drohocki (cited in **70**) found, upon repeated examination of epileptics beginning immediately after seizure, an extremely dilated picture, with evidently a number of color responses, the majority of which were unbalanced in favor of $CF+C$. Stainbrook (**82**) found somewhat similar occurrences however, with first the appearance of *Cn*, then *CF* and *FC*.

The findings of Drohocki and of Stainbrook were emphasized. Could they not suggest something such as the following? It is reasonable to suppose that, immediately following a convulsion, the synaptic connections within the brain are weakened, distended. A virtually physical increase in ego permeability would thus occur. At the least, one must admit that individuals having just experienced a severe convulsion would be dazed and would have less control of affective drives; would of necessity respond to the environment and perceive it in a rather gross, undifferentiated fashion.

Brain-injured individuals. Where actual destruction of brain tissue is known to have occurred, most investigators (**53, 64, 70, 78, 91**) find that on the Rorschach test "organics" present a picture of extratensiveness, with $CF+C$ responses predominating over *FC*. There is an additional indication that may be present where the other is not—color naming. The functions that color naming may play in the dynamics of the child were previously discussed: color naming may be a rather concrete response evoked by the practice of parents teaching their children the names of various colors by pointing to an object and simply giving the name of the surface color; or it might be due to an attempt on the part of the child to designate an emotional situation without being able to act on his interpretation. It is doubtful which, if either, of these interpretations is applicable to the brain-injured individual. If it is the former, this is an indication of the adoption of an ineffective concrete attitude. If it is the latter, a bewilderment, an inability to integrate and resolve affective situations, may be indicated.

A further contribution. Bychowski (**22**) has published a paper resulting from his observations and study of individuals undergoing treatment and training in a rehabilitation center

for the brain injured. He found that, while certain physiological changes resulting from the injury might not be alleviated, a considerable improvement in the psychological behavior of the individual frequently resulted from the retraining given by the clinic. It is as if the adaptive and emotional skills of the individual had been destroyed or eliminated at least temporarily by the injury. It was necessary for the individual to relearn, under very careful guidance and supervision, adaptive and emotional skills. The writer feels that Bychowski's report tends to confirm the postulates made in other sections of this paper about the relation between the affective picture presented by the individual and his capacity and opportunity to learn.

A Free Behavior Situation

Young and Higginbotham (**96**) attempted to correlate certain Rorschach factors with actual behavior in a free situation (summer camp). While a number of records corresponded approximately to the behavior in which the boys indulged, there were some striking exceptions —such as the child who was seemingly the most excitable and impulsive in the entire camp. This child had given no color responses on the Rorschach test, and the authors were inclined to question whether color was actually an indication of impulsivity; if it were, one might certainly expect this child to use a great deal of color, with most of it being of the *CF* or *C* variety. The authors' point was well taken. The present writer has expressed the opinion that impulsivity is not necessarily a concomitant of unbalanced color. Schachtel (**77**) suggested that affects (and color) are conscious manifestations of instinctual and other drives. An obvious conclusion in the case of the particular individual in question is that he has repressed his affective drives and is not even attempting to handle them on a conscious basis or level, or that he has never learned to recognize affective behavior in himself or others. From such an individual one would expect quite flighty and "impulsive" actions.

The Strongest Challenge

Siipola, in a highly original experiment concerned with the effect of color upon the responses of individuals taking the Rorschach test, concluded: "Apparently, the mere presence of color in a blot does not endow it with magic affect-arousing properties" (**81**, p. 381). She did find, however, an increase in the number of emotional attitudes when the individual was confronted by the colored blots and, as a result of the color, "a weak selective influence among form dominated concepts, and a strong disruptive influence involving symptoms suggestive of conceptual conflict and behavioral disorganization" (**81**, p. 381).

Many individuals when confronted with an achromatic version of the blot gave exactly the same response given by those individuals confronted with the chromatic version. One example of such behavior was the response of "butterfly" to the middle red area of Card III. The present writer has found time and time again that some individuals when confronted with this blot in the chromatic version will give the response of butterfly, and then when asked what contributed to the forming of the concept answer that it was only the form, just the shape. The color was actually ignored. There are many other individuals who do respond to the color of the blot and use the color in developing their con-

cepts. It is the *response* to color which is important, and not just the fact that the color is there to be responded to by any except the color-blind individual.

Siipola (**81**) also found that those blots which especially aroused emotional responses or emotionally charged reactions were those in which the color was particularly incongruous to the usual forms that were conceived in the achromatic version. She felt that it was this incongruity of color and form which aroused the emotional reaction rather than simply the presence of color, for where there were no incongruities the emotional reactions did not occur. This finding ties in closely with her certainly well-grounded statement that as yet no one has succeeded in bridging the gap between color and affect by other than empirical data or theory based upon empirical data.

This last finding, that of emotional reactions being the product of color and form incongruity, may force a reconceptualization of the problem. Let it be granted, for the moment, that color in and of itself has no effect. Its mere presence is of no importance. However, some individuals will respond to the colored inkblot by incorporating the color into a concept. There are other individuals who will look at this same colored inkblot without incorporating the color into a concept. The point that affect-color theorists and empiricists wish to make is that the individual who does respond to color is in some way different from the individual who does not do so. Further, these theorists and empiricists feel that the manner in which the color is used is a very keen tool for analyzing the individual's affective life, and for gaining some insight into his ego functioning. At the present time, more information exists upon the latter phase (as this review shows) than the former. A suggestion is that those individuals who do respond to color show more permeability, more susceptibility to influence by the environmental configuration than those individuals responding little or not at all to color.

Can it be granted that the mere presence of color has no effect upon the organism? Goldstein's well-known experiment with brain-injured individuals (**37**) seems to indicate clearly that exposing color to such individuals may markedly influence their physiological functioning. A study by Baccino (**7**) rather definitely indicates that chromatic illumination has a rather profound influence on the physiological functioning and growth of certain animal organisms. Conversely, the suggestion is strong that this strong response to color is actually due to definite brain damage resulting in decreasing effectiveness of inhibitory centers (**91**).

An experiment by Kravkov is highly provocative. This investigator injected into humans certain drugs which were known to have an effect upon the autonomic nervous system. He then compared the sensitivity of the eye to various colors.

"Changes in color sensation indicate a definite regularity depending upon the portion of the vegetative nervous system which is chiefly stimulated. Thus (use of) sympathetic toxins . . . bring about an increase in color sensitivity with respect to the green-blue rays of the spectrum and in contrast lower the color sensitivity with respect to the orange-red rays. The utilization of parasympathetic toxins brings about an increase in the sensitivity to orange-red rays and lowers the sensitivity to green-blue rays" (**56**, p. 94; translated from the Russian).

While these particular findings are not of tremendous importance here, the writer feels that the fact that alterations in the autonomic nervous system can cause a change in the behavior of individuals towards color is of great importance. Recently, there has been accumulating a considerable body of evidence to suggest that the condition of the autonomic nervous system may have profound effect upon behavior (**31, 89**).

A Return to the Theory

There has been sketched in this paper a theory of the relation of the response to color and personality dynamics. Two papers, one by Schachtel (**77**), the other by Rickers-Ovsiankina (**74**), contributed heavily to its formulation. The theory with some of its ramifications is briefly stated below.

The response to color is indicative of a certain functioning of the ego: the relation of the ego to its external environment, its degree of communication with, and readiness to respond to it. One may determine much concerning the affective life of an individual through an analysis of his response to color: his control of affective charges, his capacity to interpret and integrate the affective behavior of others. But Rickers-Ovsiankina (**74**) speaks of the permeability of the ego. Does this permeability refer solely to affect or could it also be extended to cover intellective functions as well? Is there actually a rigid demarcation between an inividual's affective and intellective life areas? Can an individual be shut off affectively from his environment and still participate intellectively with it?

One of the personality patterns presented by Beck (**12**) may shed illumination. Beck draws a picture of a university president, a skilled scientist who has made valuable contributions in many areas. This individual gave nine *FC* and five *CF* responses. Interpreted according to the theory presented here, this individual is exceptionally responsive to the environment in which he lives, but he also has tremendous capacity and power to integrate and control this responsiveness.

Whence comes the material with which we create but the environment? If the degree of communication with the environment is limited, then the material with which to create is limited. An individual who has limited communication with his environment may be able to do much with what he has. But does not the individual with considerable environmental communication, granted the capacity to control and integrate his responsiveness, have a tremendous advantage over his fellowman who has not this responsiveness?

A second ramification concerns the clinician or other investigator seeking to learn of the affective life of an individual and his environmental responsiveness. A point made by Schachtel (**77**) must be continually kept in mind. Affect, and color, is the conscious manifestation of instinctual and other drives. If the individual has no or little experience of affect or of color, the possible inferences are three: (*a*) he may have an inherently limited capacity for affective experience, (*b*) he may be repressing affective experience, or (*c*) he may never have learned to express his affective drives. If the second condition exists, then the individual is not dealing with his affective life on a conscious level, and one should not expect to find representations of affective life, such as the use of color, in tests such as the Rorschach and in similar situations. This condition may have been operative in the case

of the "impulsive" individual described by Young and Higginbotham (96). If the third condition exists, one should anticipate evidence of limited and perhaps stereotyped affective experience. This third condition the writer feels to be particularly exemplified by the juvenile delinquent (69), although the second condition could also be operative. The selection by the clinician of the particular condition which is in effect must rest upon an analysis of other factors and a careful case history.

A caution must be made. For the affect-color theory to be functionally correct, there need not be any emotional or affective reaction to color. It is the *response* to color which is important. There need be no "magic affect-arousing properties" of color. There may be affect-arousing properties and emotional or affective reactions to color may occur, but these factors are not ingredients of this theory.

Summary

1. A theory concerning the nature of the relation of the response to color and personality dynamics was presented. The theory strongly suggests that much can be learned from the response to color by the individual concerning the nature of the relation of the ego to the external environment as well as the relation of the ego to the affective drives of the individual.

2. A number of studies, drawn principally from the large body of Rorschach data, but also including several based upon the Mosaic Test and easel painting, were reviewed. The writer feels, first, that the theory is substantiated by the articles reviewed, and second, that a clarification of the dynamics of certain normal as well as disease processes results when the theory presented here is used in interpretation rather than prevalent practices.

BIBLIOGRAPHY

1. Abel, T. M. The Rorschach test and school success among mental defectives. *Rorsch. Res. Exch.*, 1945, **9**, 105–110.
2. Aita, J. A., *et al.* Rorschach's test as a diagnostic aid to brain injury. *Amer. J. Psychiat.*, 1947, **103**, 770–780.
3. Alschuler, R., & Hattwick, L. *Painting and personality; a study of young children.* Chicago: Univer. of Chicago Press, 1947.
4. Anastasi, Anne. An experimental study of the drawing behavior of adult psychotics in comparison with that of a normal control group. *J. exp. Psychol.*, 1944, **34**, 169–194.
5. Arlow, J., & Kadis, A. Fingerpainting in the psychotherapy of children. *Amer. J. Orthopsychiat.*, 1946, **16**, 134–146.
6. Arluck, E. W. A study of some personality differences between epileptics and normals. *Rorsch. Res. Exch.*, 1940, **4**, 154–156.
7. Baccino, M. Étude de l'influence de la lumière colorée sur la croisance des jeunes homéothermes (lapereaux, cobayes, et rats blanc). *Séance du 25 Juin, 1938. Comptes rendus Hebdomadanes de la societé d*e *biologie et de ses filiales,* 1938, **11**, 767.
8. Baker, L., & Harris, J. The validation of Rorschach test results against laboratory behavior. *J. clin. Psychol.*, 1949, **5**, 161–164.
9. Beck, S. Personality structure in schizophrenics. *Nerv. ment. Dis. Monogr.*, 1938, No. 63.
10. Beck, S. The Rorschach test in psychotherapy. *J. consult. Psychol.*, 1943, **7**, 103–111.
11. Beck, S. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune and Stratton, 1944.
12. Beck, S. *Rorschach's test.* Vol. II. *A variety of personality pictures.* New York: Grune and Stratton, 1947.
13. Beck, S., *et. al.* The normal personality as projected by the Rorschach test. *J. Psychol.*, 1950, **30**, 241–298.
14. Billig, O., & Sullivan, D. Prognostic

data in chronic alcoholism. *Rorsch. Res. Exch.*, 1942, **6**, 117–127.

15. Blum, L., & Dragowitz, A. Fingerpainting: the developmental aspects. *Child Develpm.*, 1947, **18**, 88–105.
16. Bochner, R., & Halpern, F. *The clinical application of the Rorschach test.* New York: Grune and Stratton, 1942.
17. Bowlus, D., Shotwell, A. A Rorschach of study psychopathic delinquency. *Amer. J. ment. Def.*, 1947, **52**, 23–30.
18. Boynton, P., & Walsworth, B. Emotionality test scores of delinquent and non-delinquent girls. *J. abnorm. soc. Psychol.*, 1943, **38**, 87–92.
19. Brenman, Margaret, & Reichard, S. Use of the Rorschach test in the prediction of hypnotizability. *Bull. Menninger Clin.*, 1943, **7**, 183–187.
20. Brussel, J., *et al.* The Rorschach method and post concussion syndrome. *Psychiat. Quart.*, 1942, **16**, 707–743.
21. Buhler, Charlotte. The concept of integration and the Rorschach test as a measurement of personality integration. *J. proj. Tech.*, 1950, **14**, 315–319.
22. Bychowski, G. The ego of the brain-wounded. *Psychoanal. Rev.*, 1949, **36**, 333–343.
23. Clark, J. Some MMPI correlates of color responses in the Rorschach. *J. consult. Psychol.*, 1948, **12**, 384–386.
24. Colm, H. The value of projective methods in the psychological examination of children. The Mosaic test in conjunction with the Rorschach and Binet tests. *Rorsch. Res. Exch.*, 1948, **12**, 216–237.
25. Davidson, H., & Klopfer, B. Rorschach statistics: I. Mentally retarded, normal and superior adults. *Rorsch. Res. Exch.*, 1938, **2**, 164–169.
26. Diamond, B., & Schmale, H. The Mosaic test. I. An evaluation of its clinical application. *Amer. J. Orthopsychiat.*, 1944, **14**, 237–251.
27. Elliot, M., & Merrill, F. *Social disorganization.* New York: Harper, 1941.
28. Endacott, J. L. The results of 100 male juvenile delinquents on the Rorschach ink-blot test. *J. crim. Psychopathol.*, 1941, **3**, 41–50.
29. Epstein, H., & Schwartz, A. Psychodiagnostic testing in group work (Rorschach and painting analysis technique). *Rorsch. Res. Exch.*, 1947, **11**, 23–41.
30. Ford, M. The application of the Rorschach test to young children. *Univer. Minn. Child Welf. Monogr.*, 1946, No. 23.
31. Gellhorn, E. *Autonomic regulations: their significance for physiology, psychology and neuropsychiatry.* New York: Interscience Publishers, 1942.
32. Gesell, A. L., & Ilg, F. *Infant and child in the culture of today.* New York: Harper, 1943.
33. Goldfarb, W. Personality trends in a group of enuretic children below the age of 10. *Rorsch. Res. Exch.*, 1942, **6**, 28–36.
34. Goldfarb, W. Effects of early institutional care on adolescent personality: Rorschach data. *Amer. J. Orthopsychiat.*, 1944, **14**, 441–447.
35. Goldfarb, W. Rorschach test differences between family-reared and institution-reared, and schizophrenic children. *Amer. J. Orthopsychiat.*, 1949, **19**, 624–633.
36. Goldfarb, W., & Klopfer, B. Rorschach characteristics of "institution" children. *Rorsch. Res. Exch.*, 1944, **8**, 92–100.
37. Goldstein, K. *The organism.* New York: American Book Co., 1939.
38. Guirdham, A. The Rorschach test in epileptics. *J. ment. Sci.*, 1935, **81**, 870–893.
39. Hamlin, R., *et al.* Objective Rorschach signs for groups of normal, maladjusted and neuropsychiatric subjects. *J. consult. Psychol.*, 1950, **14**, 276–282.
40. Hartman, H., *et al.* Comments on the formation of psychic structure. *Psychoanal. Stud. Child*, 1946, **2**, 11–38.
41. Hertz, M. R. Personality patterns in adolescence as portrayed by the Rorschach ink-blot method: III. The "Erlebnistypus" (a normative study). *J. gen. Psychol.*, 1943, **28**, 225–276.
42. Hertz, M. R., & Baker, E. Personality patterns in adolescence as portrayed by the Rorschach ink-blot method: II. The color factors. *J. gen. Psychol.*, 1943, **28**, 3–61.
43. Hertzman, M., & Marguiles, H. Developmental changes as reflected in Rorschach test responses. *J. genet. Psychol.*, 1943, **62**, 189–215.
44. Heuser, K. D. The psychopathic personality: Rorschach patterns of 28 cases. *Amer. J. Psychiat.*, 1946, **103**, 105–112.
45. Jersild, A. Emotional development. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1946.
46. Kates, S. Objective Rorschach response patterns differentiating anxiety reactions from obsessive-compulsive reac-

tions. *J. consult. Psychol.*, 1950, **14**, 226–229.
47. Kelley, D. M. Intravenous sodium amytal medication as an aid to the Rorschach method. *Psychiat. Quart.*, 1941, **15**, 68.
48. Kelley, D. M., & Barrera, S. The Rorschach method in the study of mental deficiency. *Amer. J. ment. Def.*, 1941, **45**, 401–407.
49. Kelley, D. M., & Klopfer, B. Application of the Rorschach method to research in schizophrenia. *Rorsch. Res. Exch.*, 1938, **3**, 55–65.
50. Kelley, D. M., & Marguiles, H. Rorschach case studies in the convulsive states. *Rorsch. Res. Exch.*, 1940, **4**, 157–189.
51. Kendig, I. Rorschach indications for the diagnosis of schizophrenia. *J. proj. Tech.*, 1949, **13**, 142–149.
52. Kisker, G. A projective approach to personality patterns during insulin-shock and metrazol therapy. *J. abnorm. soc. Psychol.*, 1942, **37**, 120.
53. Klebanoff, I. Psychological changes in organic brain lesions and ablations. *Psychol. Bull.*, 1945, **42**, 585–623.
54. Klopfer, B., & Kelley, D. *The Rorschach technique: a manual for a projective method of personality diagnosis.* Yonkers-On-Hudson, New York: World Book Co., 1946.
55. Klopfer, B., & Marguiles, H. Rorschach reactions in early childhood. *Rorsch. Res. Exch.*, 1941, **5**, 1–23.
56. Kravkov, S. O sviaziakh tsvetnogo zreniia s vegetativnoi nervnoi sistenoi. *Probl. physiol. Opt., Acad. Sci., U.S.S.R.*, 1941, **1**, 87–95.
57. Levine, K., *et al.* Hypnotically induced mood changes in the verbal and graphic Rorschach. A case study. *Rorsch. Res. Exch.*, 1944, **8**, 104–124.
58. Lindner, R. The Rorschach test and the diagnosis of psychopathic personality. *J. crim. Psychopathol.*, 1943, **5**, 69–93.
59. Linn, L. The Rorschach test results in the evaluation of military personnel. *Rorsch. Res. Exch.*, 1946, **10**, 20–27.
60. Lowenfield, M. The Mosaic test. *Amer. J. Orthopsychiat.*, 1949, **19**, 537–550.
61. McCulloch, T., & Girdner, J. Use of the Lowenfield Mosaic test with mental defectives. *Amer. J. ment. Def.*, 1949, **53**, 486–496.
62. McFate, M., & Orr, F. Through adolescence with the Rorschach. *Rorsch. Res. Exch.*, 1949, **13**, 302–319.
63. McLeod, H. A Rorschach study with pre-school children. *J. proj. Tech.*, 1950, **14**, 453–464.
64. Nadel, A. A qualitative analysis of behavior following cerebral lesions diagnosed as primarily affecting the frontal lobes. *Arch. Psychol.*, 1938, **32**, No. 224.
65. Napoli, P. Fingerpainting and personality diagnosis. *Genet. Psychol. Monogr.*, 1946, **34**, 129–230.
66. Napoli, P. Interpretative aspects of fingerpainting. *J. Psychol.*, 1947, **23**, 93–132.
67. Oeser, O. Some experiments on the abstraction of form and color: II. Rorschach tests. *Brit. J. Psychol.*, 1932, **22**, 287–323.
68. Pemberton, W. General semantics and the Rorschach test. *Papers from the Second American Congress on general semantics.* Chicago, 1943.
69. Phillips, E., & Stromberg, E. A comparative study of fingerpainting performance in detention homes and high school pupils. *J. Psychol.*, 1948, **26**, 507–515.
70. Piotrowski, Z. The Rorschach ink-blot method in organic disturbances of the central nervous system. *J. nerv. ment. Dis.*, 1937, **86**, 525–537.
71. Prados, M., & Fried, E. Personality structures of the older age groups. *J. clin. Psychol.*, 1947, **3**, 114–120.
72. Precker, J. A. Painting and drawing in personality assessment. *J. proj. Tech.*, 1950, **14**, 262–286.
73. Rickers-Ovsiankina, Maria. The Rorschach test as applied to normal and schizophrenic subjects. *Brit. J. med. Psychol.*, 1938, **17**, 227–257.
74. Rickers-Ovsiankina, Maria. Some theoretical considerations regarding the Rorschach method. *Rorsch. Res. Exch.*, 1943, **7**, 41–53.
75. Rorschach, H. *Psychodiagnostics.* New York: Grune and Stratton, 1949.
76. Sarason, S., & Potter, E. Color in the Rorschach and Kohs block designs. *J. consult. Psychol.*, 1947, **11**, 202–206.
77. Schachtel, E. On color and affect; contributions to an understanding of Rorschach's test. II. *Psychiatry*, 1943, **6**, 393–409.
78. Schafer, R. *The clinical application of psychological tests.* New York: International Universities Press, 1949.
79. Schmidl, F. The Rorschach test in delinquency research. *Amer. J. Orthopsychiat.*, 1947, **17**, 151–160.
80. Siegel, M. The diagnostic and prognostic validity of the Rorschach test in a

child guidance clinic. *Amer. J. Orthopsychiat.*, 1948, **18**, 119–133.
81. SIIPOLA, E. The influence of color on reactions to ink blots. *J. Pers.*, 1950, **18**, 358–382.
82. STAINBROOK, E. The Rorschach description of immediate post-convulsive mental function. *Charact. & Pers.*, 1943–1944, **12**, 302–322.
83. STEIN, M. Personality factors involved in the temporal development of Rorschach responses. *Rorsch. Res. Exch.*, 1949, **13**, 355–414.
84. STEINZOR, B. Rorschach responses of achieving and non-achieving college students of high ability. *Amer. J. Orthopsychiat.*, 1944, **14**, 494–504.
85. STERN, K., & MALLOY, H. Rorschach studies on patients with paranoid features. *J. clin. Psychol.*, 1945, **1**, 272–280.
86. SWIFT, J. Rorschach responses of 82 preschool children. *Rorsch. Res. Exch.*, 1945, **9**, 74–84.
87. WAEHNER, T. Interpretation of spontaneous drawings and paintings. *Genet. Psychol. Monogr.*, 1946, **33**, 3–70.
88. WARREN, H. *Dictionary of psychology.* Boston: Houghton Mifflin, 1934.
89. WENGER, M. Preliminary study of the significance of measures of autonomic balance. *Psychosom. Med.*, 1947, **9**, 301–309.
90. WERNER, H. Rorschach method applied to two clinical groups of mental defectives. *Amer. J. ment. Def.*, 1945, **49**, 304–306.
91. WERNER, H. Perceptual behavior of brain injured children: an experimental study by means of the Rorschach test. *Genet. Psychol. Monogr.*, 1945, **31**, 51–110.
92. WERTHAM, F., & GOLDEN, L. A differential diagnostic method of interpreting the Mosaic and colored block designs. *Amer. J. Orthopsychiat.*, 1941, **98**, 124–131.
93. WILKINS, W., & ADAMS, A. The use of the Rorschach test under hypnosis and under sodium amytal in military psychology. *J. gen. Psychol.*, 1947, **36**, 131–138.
94. WILLIAMS, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21–29.
95. WITTENBORN, J. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950, **14**, 261–267.
96. YOUNG, R., & HIGGINBOTHAM, S. Behavior checks on the Rorschach method. *Amer. J. Orthopsychiat.*, 1942, **12**, 87–94.

Received February 4, 1952.

Bauer, Raymond A. *The new man in Soviet psychology.* Foreword by Jerome S. Bruner. Cambridge, Mass.: Harvard Univer. Press, 1952. Pp. xxiii+229. $4.00.

In view of the very real curiosity—to use a mild term—which most of us feel regarding events on the other side of the semipermeable (?) membrane which separates us from the Russians, this little volume should be warmly welcomed by American psychologists. Since so few of us have direct access, for various reasons, to the Soviet psychological literature, and since contacts with our Russian colleagues have been reduced to the vanishing point, we are all indebted to Bauer for bringing us relatively up to date on recent developments.

From the point of view of scientific psychology the picture which Bauer presents is generally discouraging. No journal, specifically psychological in nature, has appeared since 1934; articles of psychological interest are published mainly in a journal devoted to pedagogy. There has been a strong reaction, still evident, against psychology as an independent discipline. The study of attitudes has been condemned and virtually abandoned. No public opinion surveys may be conducted, and social psychology in general "has become virtually a proscribed area" (p. 169). A Party decree in 1936 resulted in the almost complete suppression of the use of psychological tests, which were "formally characterized as instruments for perpetuating the class structure of bourgeois societies" (p. 124). Scientific theory, in psychology as elsewhere, is validated not in terms of its relation to empirically verified facts, but by the contribution it can make to the Party's program. A psychologist is quoted as stating that "every theoretical mistake, every error in the field of methodology is inescapably transferred into a political error" (p. 106). It goes without saying that "incorrect" views cannot be expressed or tolerated.

Bauer characterizes his book as "partially a history of the science of psychology in the Soviet Union, partially a study of the pattern of social change in that country, largely an analysis of changing conceptions of human nature under conditions of social change, to a certain extent an inquiry in the relation of ideology to action, somewhat a study of the relationship of psychology to society" (p. ix). On the whole this ambitious program is effectively realized. There is, of course, a close relation between the character and structure of a society and the current beliefs concerning the nature of man, his goals and motives, his development and socialization. Bauer has presented some striking correspondences between the character of the society as a whole, and the specific developments that have occurred in the field of psychology.

Nowhere does this come out more clearly than in his discussion of the changes which have occurred in the concept of the nature of man, as a reflection of the changes which took place in the political scene. In the 1920's, for example, Soviet psychologists proceeded on the assumption that man's nature was essentially passive, his characteristics determined by the (mainly economic) environment. Fundamentally good and noble, man had been misled and perverted by the evil (that is, bourgeois,

capitalist) system under which he lived. Even after the Bolshevik revolution, the environment could still be held responsible, because it contained many of the elements of the older social and economic structure. By 1936, however, socialism had allegedly been fully realized in the Soviet Union. From that time on, it became impossible to blame the (socialist) environment; the responsibility was now placed upon the individual himself. "The dominant conception of man became that of an increasingly purposeful being, who was more and more the master of his own fate, and less and less the creature of his environment" (p. 7).

That meant a movement away from behaviorism, reflexology, and "reactology" to a more "purposive" variety of psychology. Consciousness was restored to a dominant role in human affairs, and the unconscious fell correspondingly into disfavor; not that its existence was denied, but rather that it became subordinate in importance to conscious, purposive action. The source of error was now found at least in part in man himself; he needed the right training, and the right self-training, to set him upon the proper road. From the point of view of Western psychology, training is of course a part of what we would call the environment. Bauer suggests that this distinction is drawn by Soviet psychologists "mainly to deprecate the importance of such aspects of the environment as the actual material conditions under which the child lives" (p. 148). The question arises as to whether the words for "environment" have a somewhat different connotation in Russian and English respectively, in which case the Soviet attack on environmental explanations of behavior would really represent an attack on a very restricted variety of environmentalism. In any case, the coincidence between the political pronouncements of 1936 and the attack of the prevailing science of psychology is a striking one.

Bauer draws an interesting contrast between the Soviet view of man and that prevalent in Nazi Germany. The Nazi view held that man was moved primarily by the unconscious, the nonrational; the Bolshevik stresses consciousness and rationality. The Nazi stressed man's weakness, his helplessness, his need of a leader; the Bolsheviks insist on man's responsibility for his behavior, on his ability to make his own destiny—though the only "right" destiny is that which follows the party line. "For the Nazi, man was a marionette who moved when one pulled the strings. For the Bolshevik, he is a robot who can be trained to act independently within specified limits" (p. 178). One may argue about some of the details of these characterizations, but it seems clear that the two dictatorships do differ markedly in the meanings they attach to human nature. Perhaps we have here a clue to a differential diagnosis of varieties of dictatorship in psychological terms.

Bauer's informative study would have been still more valuable, at least in the opinion of this reviewer, if he had included somewhat more discussion of some of the specific investigations carried out by Russian psychologists. Granting that his interest was in the development of theory, his thesis could have been more clearly illuminated by a fuller demonstration of the manner and extent to which theory dominated the collection of "facts." One might argue also about the amount of emphasis which Bauer places on the discontinuities to be found in Soviet psychology. It is interesting that Dr. Joseph Wortis in his *Soviet Psychia-*

try (Baltimore, 1950), as Bauer himself indicates, was much more forcibly struck by the continuities. This difference in interpretation by two scholars examining closely related material calls for fuller exploration than that contained in a brief footnote. These are relatively minor issues, however, compared to the over-all value of Bauer's study. The Russian Research Center at Harvard has made a significant addition to its excellent series.

The last sentence of the book is worth repeating. "Political interference in science does not destroy completely the usefulness of science to the system, but the continued suppression of freedom of scientific inquiry must ultimately lead to the point where the society cannot solve its own problems effectively" (p. 196). This applies in the United States just as it does elsewhere. It cannot be said too often: there can be no real development of science where there is no freedom—freedom to explore, to doubt, to criticize, to deviate, even to be wrong. There are people in this country, too, who have taken it upon themselves to tell scientists, psychologists and others, what they may teach and what they may discover. Soviet psychology should serve us as an object lesson. If we allow that kind of interference here, we might just as well shut up shop.

Otto Klineberg.
Columbia University.

Jaques, Elliott. *The changing culture of a factory*. New York: Dryden Press, 1952. Pp. xxi+341. $4.25.

This is the published report of ". . . a case study of developments in the social life of one industrial community between April, 1948 and November 1950." The "case" is a small, publicly held British company engaged principally in the manufacture, sale, and servicing of metal bearings. The study is concerned with the description, diagnosis, and treatment of the corporate syntality. The results reported are the product of the collaborative efforts of the personnel of the company and of a thirteen-member research team headed by the author of the book, Dr. Jaques. The research was sponsored by the Tavistock Institute of Human Relations, London, and has been accepted as a Ph.D. thesis in the Department of Social Relations at Harvard.

The first part provides a retrospective glimpse of the corporate organization as it evolved during the first fifty years of life. This is followed by the case study proper, a detailed description of events as they occurred during the period of observation. The methods used by the research team to gain acceptance by company personnel at all levels and to function effectively in the multiple role of consultant, analyst, and therapist are a highlight of this second section of the report. The third and concluding part of the book contains an analysis and interpretation of the findings. "The method of analysis . . . will be to study how the pattern of social activity at Glacier (firm name) . . . has come about through the interaction of the firm's organizational structure, its customary way of doing things, and the behavior of its members . . . we shall study the interaction of social structure, culture, and personality." The results presented in this part point up the need for defining and clarifying individual and group roles as an antecedent step both to understanding social behavior and to evaluating it. Inferentially, the adequacy of a group's adjustment, in large measure, is considered to be a function of the members' understanding of the au-

thority-responsibility relationship and possession of authority by the individual members commensurate with their felt responsibilities.

The study exemplifies social science at its best, transcending the boundaries of any single professional research area. From the standpoint of both content and methodology it warrants the attention of all psychologists concerned with interpersonal and intergroup relations as they affect the individual's adjustment. Generalization from the specific findings is, of course, limited by the very nature of the case study method and, in this instance, by the concurrent diagnosis and counseling required of the investigators during the course of the study.

WILLIAM J. E. CRISSY.
Queens College.

JUDD, DEANE B. *Color in business, science and industry.* New York: Wiley, 1952. Pp. ix+401. $6.50.

The author of *Color in Business, Science and Industry* seems to address himself primarily to business men and industrialists to call attention to the scientific aids now available for the solution of a variety of practical color problems. In the preface he says, "It has been my privilege . . . in my twenty years at the National Bureau of Standards, to come into contact with hundreds of colorimetric sore spots in our industrial life. I have seen victories that paid off in dollars and cents won by applying the sciences of mathematics, physics, and psychology to these problems." The author refers specifically to a great number of practical color problems encountered in everyday life and endeavors to indicate how "visual psychophysics mixed with a liberal sprinkling of common sense" can provide a solution for these problems.

The work is divided into three principal sections: Part I, Basic Facts; Part II, Tools and Technics; and Part III, Physics and Psychophysics of Colorant Layers. Cursory inspection reveals that Part II is the most important. More than half the book is devoted to this part, which is nearly three times as long as Part I and four times as long as Part III.

Part I lays the groundwork for the later exposition. It includes a twenty-page treatment on the structure and functions of the eye, a summary of the basic physical, psychological, and psychophysical terms currently employed in the field of color perception, a discussion of methods of color matching (*a*) by addition of lights, (*b*) by rapid succession of lights, and (*c*) by mixture of colorants. The first part closes with a discussion of different types of color deficiency and a brief description of the better known tests of color blindness.

The various tools and technics used by the color specialist are described in Part II. The reader is informed in some detail, with the aid of many diagrams and with the necessary quantitative tables, concerning spectrophotometers, the standard observer, chromaticity diagrams, tristimulus values and tristimulus colorimeters, subtractive colorimeters, photometers, photoelectric tristimulus colorimeters, color standards, color scales, and color names. How each of these aids is to be applied in connection with diverse manufacturing problems is set forth in an easy running style with great clarity. Here the author has rendered important service to workers in the color practicum, not only by indicating what each of these aids is designed to do and how it is to be applied, but also by pointing out some of their limitations.

Part III is devoted to special problems, such as gloss, opacity or hiding power, clear and turbid media, which

have been subjected to intensive quantitative analyses. The solutions of these problems, sometimes involving exponential and hyperbolic functions, will be of interest chiefly to those of high mathematical competence.

About fifteen pages are devoted to each of the three final sections of the work: (*a*) an appendix, containing quantitative tables too extended to be included in the main body of the text, (*b*) a list of references, and (*c*) the index. Although the list of references contains more than three hundred and fifty items, a number of very important contributors to color are omitted. There is no reference to E. Hering, G. E. Müller, L. T. Troland, C. E. Ferree, and S. Hecht, to mention only a few of those no longer living.

Of the several sciences sharing in the scientific study of color, physics and psychology fare somewhat better than physiology. However, the opening treatment in Part I is designed to provide some balance among the several sciences with overlapping interests in color. Despite the attempted integration among these sciences, it seems fair to say that Judd's chief contribution consists in the description of the methods by which the physical correlates of the visual stimulus are to be specified. What seems to be left for other men of science, perhaps in electrophysiology, is the task of surveying more fully the possibilities of specifying and, in so far as possible, of rendering to quantitative terms the physiological correlates of the visual stimulus.

MICHAEL J. ZIGLER.
Wellesley College.

KARN, HARRY W., AND GILMER, B. VON HALLER. *Readings in industrial and business psychology.* New York: McGraw-Hill, 1952. Pp. ix+476. $4.50.

The trend toward books of readings in various areas of psychology is continued with the publication of this volume. Some such books have tended to present materials from all eras of psychology, some have had the original articles edited rather sharply, and some have included extensive comments by the editors directed toward emphasis and integration. The present volume does none of these. The editors state that they have not attempted integration nor have they tried to cover articles of historical interest. They have tried to include "representative" articles which are easily understood by those lacking extensive technical training.

The aim of recency is realized, 41 of the 53 articles having been published since 1944. The articles are drawn from 19 different sources with approximately 25 per cent of the articles having been published originally in either the *Journal of Applied Psychology* or *Personnel Psychology.* The articles are rather evenly spread among 11 "fields" of business and industrial psychology. There is some question regarding "representativeness" in the selected articles. For example, in terms of content, the editors might have selected a more representative article on sampling in market research than the one by Stanton written in 1941 which makes no reference to probability sampling. Furthermore, the number of articles devoted to each of the fields is not proportionate to the amount of research or information in those fields. Although this reviewer is not opposed to disproportionate sampling of articles, the term representative is hardly used in its usual sense and should be read as illustrative. It is good to see sections devoted to the place of the psychologist in industry and to his ethical problems. The

comments made by the editors about each article are extremely brief and of no great value. These comments could have been expanded and might have included a few words about each author.

As far as this reviewer knows, this is the first attempt to provide a book of general readings in business and industrial psychology since Moore and Hartmann's *Readings* of 20 years ago. The latter largely contained excerpts, whereas Karn and Gilmer present complete articles. Direct comparison of the books is probably unfair since the selections were not made on the same bases in the two books. However, a casual comparison shows that the present volume devotes more space to training, counseling, job evaluation, market research, fatigue and efficiency, and leadership, and much less to tests and selection than did the earlier book. About the same amount of space is devoted to motivation and morale, and industrial relations.

As the editors point out, this book should be a supplement to a systematic text or else the teacher must organize and integrate the subject matter. The main advantage to this book is that a number of recent, illustrative articles are combined under one cover.

LESTER GUEST.
The Pennsylvania State College.

YOUNG, KIMBALL. *Personality and problems of adjustment.* (2nd Ed.) New York: Appleton-Century-Crofts, 1952. Pp. x+716. $5.00.

A considerable amount has been added to the first edition of this very readable book, which appeared first in 1940 (see *Psychological Bulletin*, 1941, 188ff.). The author has included some fairly recent material from the perception-personality area, but the presentation is substantially eclectic. Every sort of approach has been covered, including George Mead's subjective analysis of the "I" and "Me," anthropological and psychoanalytical and field theories, and the personalistic contribution of Gordon Allport, leading into case studies. Some might object to Young's disinclination to take a point of view and stick to it. However, the average student will probably be benefited by the broad coverage.

The author is quite skeptical of our progress in the strictly experimental study of personality measurement. He states, "there are a large number of theories (of personality) but unfortunately they are not, for the most part, so stated as to furnish a bridge to empirical testing either in the laboratory or by other scientific devices."

Part I is mainly concerned with theories about personality and its development. There are traditional presentations of language and other forms of learning. Little has been added here, and the chapters on symbolic behavior and the self have apparently not been altered much. Both of these areas remain useful summaries of often neglected topics.

Part II, concerned with problems of adjustment, is the most concrete and will doubtless be read with major interest. As in the first edition, there are chapters devoted to infancy, childhood, adolescence, the college student, marriage, and neuroticism. There is in addition a completely new chapter on later maturity and old age, indicative of this rapidly expanding field. This reviewer feels that the chapter on psychological problems associated with occupation is given less up-to-date attention than it deserves, in comparison, for example, with the far greater consideration given to sexual and marital adjustment problems.

EDWARD S. JONES.
University of Buffalo.

THOMPSON, GEORGE G. *Child psychology: growth trends in psychological adjustment.* Boston: Houghton Mifflin, 1952. Pp. xxxiv+667. $5.50.

Oftentimes the child psychologist finds it difficult to convince himself or his colleagues that his area of interest has any specific contribution to make to psychology as a whole. In this book Thompson presents a convincing demonstration that when it is interpreted as the psychology of *development* (rather than as the psychology of children, to be contrasted with the psychology of apes or adolescents or old people) child psychology is a scientific discipline in its own right. This is a book developed thoughtfully, carefully, and with a high degree of scholarship. Obviously familiar in detail with the vast literature in developmental psychology, the author has been guided in his selection of material by criteria of scientific validity and pertinence of the material to the point he is trying to make. In other words, while the book is comprehensive it is not padded. Each of its fourteen chapters is organized as a unit and these units contribute to a functional whole.

Thompson's objective seems to have been to present an integrated picture of the developmental process underlying human behavior. Where necessary to round out the picture he has felt free to use findings with infrahuman "children," studies of individual cases, concepts from personality theory and the like, pointing out frequently that in many important areas of child psychology our knowledge is all too sparse. But pervading the whole is the insistence on the application of scientific principles in evaluating data. Considerations such as numbers of cases, the design of studies, and the reliability of observation are given repeated and consistent emphasis.

At various points in the book it becomes apparent that behavioral development does not always take place in a desirable direction, and at these points the author is particularly interested in examining any evidence which might explain such trends. This leads to careful consideration of methods of child rearing and guidance and education, in our own culture and in others.

In his general orientation, the author's position is not far from that of the modern behaviorist, but he has given considerable effort to the task of presenting fairly such materials as projective techniques and the theory underlying their use, if with an eclectic sort of damning with faint praise. Placing a premium upon factors of validity and reliability, it is natural that he would say of these techniques, "we should be cautious about our interpretations until more objective methods of scoring and interpretation have been worked out, and until more validation studies have been conducted with positive results" (p. 620). Exercising perhaps the same requirement of objective proof, the author gives almost no space at all to therapy as such.

This book will be well received by those who feel that child psychology should be taught as a science. It is written in a fashion sufficiently interesting to be appropriate for the non-psychology major who wants to learn something about children before generating his own, perhaps, and in a fashion sufficiently scholarly and comprehensive to satisfy the graduate student looking for a standard reference in this field. This book is to be highly recommended.

T. W. RICHARDS.
Louisiana State University.

BOOKS AND MONOGRAPHS RECEIVED

Ashby, W. Ross. *Design for a brain.* New York: Wiley, 1952. Pp. ix+259. $6.00.

Asch, Solomon. *Social psychology.* New York: Prentice-Hall, 1952. Pp. xvi+646. $5.50.

Barlow, Fred. *Mental prodigies.* New York: Philosophical Library, 1952. Pp. 256. $4.75.

Bender, Morris B. *Disorders in perception.* Springfield, Ill.: Charles C Thomas, 1952. Pp. viii+109. $3.00.

Bergler, Edmund. *The superego.* New York. Grune & Stratton, 1952. Pp. x+367. $6.75.

Bernard, Harold W. *Mental hygiene for classroom teachers.* New York McGraw-Hill, 1952. Pp. xiii+472. $4.75.

Bett, W. R. *The infirmities of genius.* New York: Philosophical Library, 1952. Pp. 192. $4.75.

Davidson, Henry A. *Forensic psychiatry.* New York: Ronald Press, 1952. Pp. viii+398. $8.00.

Deese, James. *The psychology of learning.* New York: McGraw-Hill, 1952. Pp. ix+398. $5.50.

Ferguson, Leonard W. *Personality measurement.* New York: McGraw-Hill, 1952. Pp. xv+457. $6.00.

Hirsh, Ira J. *The measurement of hearing.* New York: McGraw-Hill, 1952. Pp. ix+364. $6.00.

Jersild, Arthur T. *In search of self.* New York: Bureau of Publications, Teachers College, Columbia Univer., 1952. Pp. xii+141. $2.75.

Kerr, Madeline. *Personality and conflict in Jamaica.* Liverpool: Univer. Press of Liverpool, 1952. Pp. xii+221. 15s.

Lindner, Robert. *Prescription for rebellion.* New York: Rinehart, 1952. Pp. 305. $3.50.

Magne, Olof. *Perception and learning.* Göteborg: Gumperts Förlag, 1952. Pp. 228. 25 kronor.

Maslow, Paul. *Powers of the mind.* Vol. II of the Life Science. Brooklyn: Author, 1952. Pp. 135–288. $3.50 (Mimeo).

McCulloch, Warren S. *Finality and form.* Springfield, Ill.: Charles C Thomas, 1952. Pp. 63. $3.75.

Merton, Robert K., Fiske, Marjorie, & Kendall, Patricia. *The focused interview.* New York: Bureau of Applied Social Research, Columbia Univer., 1952. Pp. xxv+202. $3.00.

Meyer, Alan S. *Social and psychological factors in opiate addiction.* New York: Bureau of Applied Social Research, Columbia Univer., 1952. Pp. ix+170. $1.00.

Rausch, Edwin. *Struktur und Metrik figural-optischer Wahrnehmung.* Frankfort am Main: Dr. Walter Kramer, 1952. Pp. xiv+404.

Smith, Gudmund. *Interpretations of behavior sequences.* København: Society of Psychology and Pedagogy, 1952. Pp. 120. 9:50 cr.

Swanson, Guy E., Newcomb, Theodore M., & Hartley, Eugene L. *Readings in social psychology.* (Rev. Ed.) New York: Holt, 1952. Pp. xviii+680. $5.00.

von Bertalanffy, Ludwig. *Problems of life.* New York: Wiley, 1952. Pp. xi+216. $4.00.

Walls, Gordon L., & Mathews, Ravenna W. *New means of studying color blindness and normal foveal color vision.* Berkeley: Univer. of California Press, 1952. Pp. iv+172. $2.50.

Weinland, James D., & Gross, Margaret V. *Personnel interviewing.* New York: Ronald, 1952. Pp. vii+416. $6.00.

Witherington, H. Carl. *Educational psychology.* (Rev. Ed.) Boston: Ginn, 1952. Pp. v+487. $4.00.

GEORGE BANTA PUBLISHING COMPANY, MENASHA, WISCONSIN